THESIS

To obtain the degree of

DOCTOR OF AIX-MARSEILLE UNIVERSITY

Doctoral School: *Sciences de la Vie et de la Santé* (ED62)

Speciality: Genomics and Bioinformatics

Research unit: Biotechnology and Biodiversity of Fungi

Genome diversity and evolutionary routes to lignocellulose degradation in wood decay fungi

Presented by

Hayat Hage

Directed by Marie-Noëlle Rosso

March 23, 2021

Presented in front of the jury:

Pedro COUTINHO	Professeur, Aix-Marseille Université, Marseille	President
Etienne DANCHIN	Directeur de Recherche, INRAE-Université Côte d'Azur, Institut Sophia-Agrobiotech, Sophia-Antipolis	Reviewer
Florence FORGET	Directeur de Recherche, INRAE, Mycologie et Sécurité des Aliments, Bordeaux	Reviewer
Francis MARTIN	Directeur de Recherche, INRAE-Université de Lorraine, Interactions Arbres/Micro-organismes, Nancy	Examiner
Helene CHIAPELLO	Ingénieur de Recherche, INRAE-Université Paris Saclay, Mathématiques et Informatique Appliquées du Génome à l'Environnement, Paris	Examiner
Marie-Noëlle ROSSO	Directeur de Recherche, INRAE- Aix-Marseille University Biodiversité et biotechnologie fongique, Marseille	Thesis Director

Affidavit

I, undersigned, Hayat Hage, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific direction of Marie-Noëlle Rosso, in accordance with the principles of honesty, integrity and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with both the French national charter for Research Integrity and the Aix-Marseille University charter on the fight against plagiarism.

This work has not been submitted previously either in this country or in another country in the same or in a similar version to any other examination body.

Marseille, 16 décembre 2020

Résumé

Les champignons dégradeurs du bois jouent un rôle majeur dans la libération du carbone séquestré dans la matière organique des écosystèmes forestiers. Ils ont évolué depuis plus de 290 millions d'années pour s'adapter aux ressources disponibles dans leur niche écologique et à des conditions environnementales difficiles. En particulier, ils ont acquis la capacité de dégrader les polymères récalcitrants des parois végétales pour en extraire des nutriments carbonés. Malgré de nombreuses découvertes sur les mécanismes enzymatiques impliqués dans la dégradation du bois, notre vision sur l'adaptation évolutive de ces champignons et leur diversité génomique reste incomplète. Ainsi, l'objectif général de ma thèse était d'explorer leur diversité fonctionnelle et génomique à différents niveaux taxonomiques, couvrant d'abord un seul genre, *Pycnoporus* spp., puis un ordre taxonomique, les Polyporales. L'analyse comparative des génomes, transcriptomes et sécrétomes des quatre espèces de *Pycnoporus* décrites à ce jour a montré une structure du génome et des répertoires de gènes très conservés et nous a permis d'identifier un jeu commun d'enzymes mobilisées en réponse à divers substrats lignocellulosiques. L'analyse génomique comparative et phylogénomique a ensuite été étendue à 50 espèces de Polyporales, dont 26 génomes nouvellement séquencés. Nous avons identifié des expansions de familles de gènes révélant différentes trajectoires d'adaptation à la décomposition du bois chez les Polyporales, ainsi que des enzymes encore peu étudiées ayant un rôle potentiel dans la décomposition du bois. Finalement, nous avons développé un outil bioinformatique pour l'identification, dans l'ensemble du règne fongique, des enzymes impliquées dans la cyclisation des sesquiterpènes, métabolites secondaires impliqués dans la communication chimique entre micro-organismes dans leur habitat naturel. Grâce à cet outil, nous avons analysé 1420 génomes fongiques et identifié 11085 gènes candidats codant pour des sesquiterpène synthases (STS), dont 55% n'auraient pas été détectés par les outils actuellement disponibles. Nous avons exploré la distribution taxonomique des enzymes STS identifiées et prédit leurs mécanismes de cyclisation. Des essais biochimiques seront maintenant nécessaires pour évaluer la précision des prédictions fonctionnelles. Nos découvertes sur la génomique des champignons dégradeurs du bois contribueront à mieux comprendre leur rôle écologique dans les environnements naturels. En outre, les systèmes enzymatiques fongiques que nous avons identifiés pourraient être testés comme biocatalyseurs pour la production de molécules à valeur ajoutée à partir de matières végétales récalcitrantes comme sources de carbone renouvelable.

Abstract

Wood-decay fungi play a crucial role in carbon release from dead organic matter in forest ecosystems. They have evolved for over 290 million years to become most efficient in utilizing the available resources of their ecological niches and to survive in harsh environmental conditions. The survival strategy of wood decay fungi largely relies on their ability to degrade recalcitrant plant cell wall polymers, which they use as a carbon source. Despite many discoveries on the enzymatic mechanisms involved in wood decomposition, our vision on the evolutionary adaptation to wood decay and genome diversity remains incomplete. The general objective of my thesis was to explore the functional and genomic diversity among wood decayers at different taxonomical levels, covering first a single genus, *Pycnoporus*, then a taxonomic order, Polyporales. The comparative genomics, transcriptomics and secretomics analysis of the four *Pycnoporus* species described to date showed limited diversity in their genome structure and repertoires of protein coding genes. This allowed us to identify a core set of enzymes mobilized by the fungi in response to diverse lignocellulosic substrates. We further extended the comparative genomic and phylogenomic analysis to 50 species (including 26 newly sequenced genomes) that cover the order Polyporales. Interestingly, our analysis highlighted gene family expansions that supported different trajectories for adaption to wood decay in Polyporales, and detected a set of conserved, yet overlooked, proteins as new candidate players in wood decomposition. In addition, we built a bioinformatic tool to screen the whole fungal kingdom for sesquiterpene synthases (STS), which are involved in the synthesis of secondary metabolite used by the fungi to compete with other microorganisms in their natural habitat. Using this tool, we analyzed 1,420 fungal genomes and identified 11,085 candidate STS, out of which 55% would have been missed by the currently available tools. We explored the taxonomic distribution of the identified STS and predicted their mechanisms of cyclisation. But biochemical assays are now necessary to assess the accuracy of the functional predictions. Overall, our findings on the genomic features of wood decay fungi will contribute to better understanding the ecological roles of these fungi in natural environments. Furthermore, the fungal enzymatic systems we identified could be tested as biocatalysts for the production of value-added molecules using recalcitrant plant materials as a renewable carbon source.

Acknowledgments

I would like to start by expressing my sincere gratitude to my PhD mentor, Dr. Marie-Noelle Rosso. Thank you for your continuous support, encouragement, and positivity. I truly enjoyed your way of guiding, always with a smile, ready to answer my questions and find the right words to make things clear and simple. Your way of doing science, your expended network and your deep knowledge made my PhD go very smoothly. Thank you for everything you have done for me during the last 3 years, I could not wish for a better mentor.

I want to extend my thanks to Dr. Craig Faulds, the director of BBF lab. Thank you for being a very cool director, for always being there to help and for your precious advices. I enjoyed and learned a lot from all the discussions we have had and from all the books you have given me.

I would like to thank INRAE and Region PACA for funding my thesis. I would also like to thank GDR Génomique Environmentale and Institut Carnot 3BCAR for funding my stays and travel expenses to the IAM lab in Nancy and to the JGI lab in Berkeley, California respectively.

I want to thank the jury of my thesis, Prof. Pedro Coutinho, Dr. Florence Forget, Ms. Helene Chiapello, Dr. Etienne Danchin and Dr. Francis Martin, who accepted to take the time to read and evaluate my work. A special thanks to Dr. Danchin and Dr. Valentin Loux who also gave me precious advices during thesis committee and Dr. Martin for receiving me in his laboratory for a month and for his helpful suggestions all along my PhD.

I want to also thank Dr. Igor Grigoriev for receiving me for two months in his laboratory, as well as all his lab members, specially Byoungnam, Andrew and Sajeet for the daily Tesla trips to try new cuisines!

Shingo, thank you for introducing me to your amazing X-INGO tools. I enjoyed a lot using them and they were a good reference for me when I started building my own tools.

Thanks to all the BBF people for the daily chitchat and good ambiance. Those who left BBF by now: Bassem, Amani, Marianne (I loved inheriting your apart!), Majid and Aurore, thanks for all the fun moments we have spent together… Elodie, thank you for your help whenever I needed. Margot, it was nice to have all the long discussions with you about life, people and your adorable daughter Lola! Delphine, I enjoyed sharing the office with you and learning the daily tips on how to become more ecolo and how to do cute stuff with the crochet! Bastien, David and Jean-Lou, so many good memories with you, from all the good days we had in the Callanques, our boat trips, the trip to Rome, the volleyball matches and the nights we spent playing Jeux de Société… Wissal, my best PhD friend! I am so glad that I met you, you are such a fun person, always having a story to tell and always being ready to go out, party or do whatever!

Tony and Manish, thanks for the long cooking times, our infinite discussions about religions and for being the best Neighbors!

Leila, Sara and Guita, remember three years ago when I told you I am moving to Marseille, you were so angry, “Can’t you find a PhD in Nantes?!” you were saying. And now here we are 3 years later, time flies! Hopefully we will be able to meet more often now. We have shared the most memorable times together in India, and I will always be excited to meet you to remember all the funny stories!

Hanane, Iman and Lara, my childhood friends, I love how nothing change between us despite being separated with the distance; Every time we meet again, it is like we have never been apart.

Sandra and Rouba, I am so glad to have cousins like you who always want to stay updated about whatever is happening in each other’s life. Thank you for all the “EMAAR” meetings that took my mind off the thesis to our imaginary future as neighbors with dozens of kids 😊.

Finally, I want to express the warmest thanks to the most important people in my life, my family. My little brothers Ahmad and Ibrahim, for being so adorable and for always waiting for us to come back home to show us your new inventions… Julia, Mayssane and Lina, how lucky I am to have sisters like you, who always understand me and on who I can always count… My parents, I feel that the words seem ridiculous compared to all you have done for me all along my 26 years… It’s thanks to you if I finally reach here today. Thank you for your continuous support, sacrifices and unconditional love. And to end, my Rik, thank you for believing in me, sometimes more than I do; for making me understand that nothing is impossible, which motivates me to give the best of me and for making me happier by being so perfect everyday!

List of Publications

Hage H*, Miyauchi S*, Virágh M, Drula E, Min B, Chaduli D, Navarro D, Favel A, Norest M, Lesage-Meessen L, Bálint B, Merényi Z, de Eugenio L, Morin E, Martínez A, Baldrian P, Štursová M, Martínez MJ, Novotny C, Magnuson J, Spatafora J, Maurice S, Pangilinan J, Andreopoulos W, LaButti K, Hundley H, Na H, Kuo A, Barry K, Lipzen A, Henrissat B, Riley R, Ahrendt S, Nagy L, Grigoriev I, Martin F, Rosso MN. Gene family expansions and transcriptome signatures uncover fungal adaptations to wood decay. (*under review - Environmental microbiology).*

Miyauchi S*, **Hage H***, Drula E, Lesage-Meessen L, Berrin JG, Navarro D, Favel A, Chaduli D, Grisel S, Haon M, Piumi F, Levasseur A, Lomascolo A, Ahrendt S, Barry K, LaButti K, Chevret D, Daum C, Mariette J, Klopp C, Cullen D, de Vries R, Gathman A, Hainaut M, Henrissat B, Hildén K, Kües U, Lilly W, Lipzen A, Mäkelä M, Martinez A, Morel-Rouhier M, Morin E, Pangilinan J, Ram A, Wösten H, Ruiz-Dueñas F, Riley R, Record E, Grigoriev I, Rosso MN. Conserved white-rot enzymatic mechanism for wood decay in the Basidiomycota genus *Pycnoporus*, *DNA Research*, Volume 27, Issue 2, April 2020.

Hage H, Asaf S, Iacazio G, Grigoriev I, Rosso MN. An HMM approach to genome-wide identification of fungal sesquiterpene synthases. (*In preparation*).

Hage H, Rosso MN, Tarrago L. Characterization of methionine sulfoxide reductase families in fungi and evidence for horizontal gene transfers. *(In preparation*).

Miyauchi S, Rancon A, Drula E, **Hage H**, Chaduli D, Favel A, Grisel S, Henrissat B, Herpoël-Gimbert I, Ruiz-Dueñas FJ, Chevret D, Hainaut M, Lin J, Wang M, Pangilinan J, Lipzen A, Lesage-Meessen L, Navarro D, Riley R, V. Grigoriev I, Zhou S, Raouche S, Rosso MN. Integrative visual omics of the white-rot fungus *Polyporus brumalis* exposes the biotechnological potential of its oxidative enzymes for delignifying raw plant biomass. *Biotechnology for biofuels*. 2018.

Bissaro B*, Kodama S*, **Hage H**, Ribeaucourt D, Haon M, Grisel S, Simaan A.J, Forget S.M, Brumer H, Rosso MN, O´Connell R, Lafond M, Yasuyuki K and Berrin JG. Tandem fungal oxidases drive pathogenesis in major food crops. (*Submitted for publication*).

List of Abbreviations

AA: Auxiliary activity
BR : Brown-rot
CaZyme : Carbohydrate active enzyme
CBM1 : Carbohydrate binding module
CDH: Cellobiose dehydrogenase
FCW : Fungal cell wall
FPP : Farnesyl diphosphate
GP : Generic peroxidase
HGT : Horizontal gene transfer
HMM : Hidden markov model
LiP : Lignin peroxidase
LPMO : Lytic polysaccharide monooxygenase
MnP : Manganese peroxidase
MRCA : Most recent common ancestor
PCW : Plant cell wall
POD : Class II peroxidase
PPI : Inorganic pyrophosphate
RAxML : Randomized axelerated maximum likelihood
ROS : Reactive oxygen species
SOM : Self-organizing map
SSP : Small secreted protein
STS : Sesquiterpene synthase
TE : Transposable element
TPS : Terpene synthase
VP : Versatile peroxidase

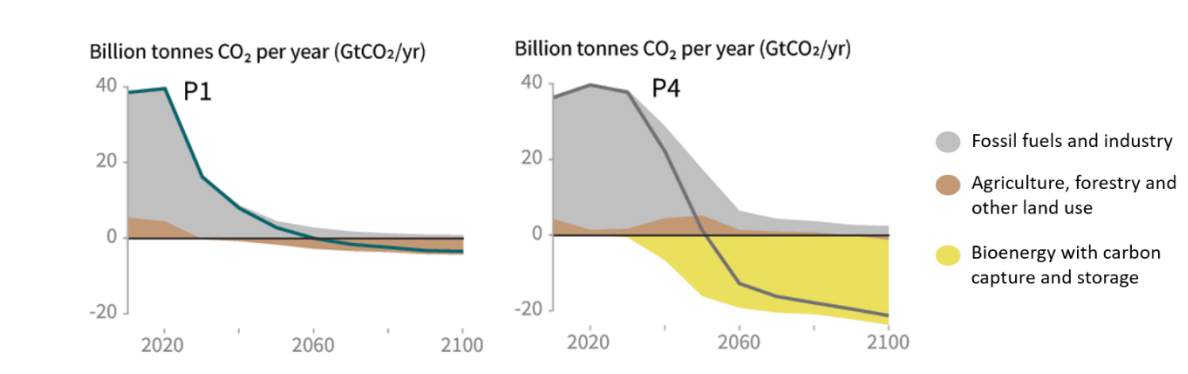
Table of Content

Chapter I CONTEXT

I.1. A bioeconomy to face the climate change

I.1.1. Urgent need for a reduction of CO_2 emissions

The massive population growth that the world faces today is causing a rise of greenhouse gas emission, among which carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O) and fluorinated gases (F-gases) are the most important. Strikingly, the CO_2 concentration in the atmosphere increased by 47% since the industrial revolution (1900s) (Buis, 2019). Meanwhile, human-induced warming reached approximately 1°C above pre-industrial levels in 2017, increasing at 0.2°C per decade (Hoegh-Guldberg *et al.*, 2018). This overload in CO_2 originates mostly from fossil fuels that are burned to fulfil human needs in energy. The accumulation of CO_2 in the atmosphere is the major cause of the rise in global temperature, which is now estimated to warm up above +1.5°C by 2050 (Hoegh-Guldberg *et al.*, 2018). Scientists and environmentalists warn that if nothing is done to deal with the global warming, catastrophic consequences may occur at the social and economic level, and the ecosystems will be unable to naturally adapt to the climate changes. In some scenarios, stabilizing the climate requires drastic reductions in energy demand while the overall living standard rises, especially in the global South. Such scenarios are dependent on technological innovations for rapid decarbonization of energy supply, low-carbon technology innovation and well-managed land systems that result in a net uptake of CO_2 from the atmosphere, for example via afforestation and reforestation, land restoration and soil carbon sequestration. In other scenarios, the energy demand keeps on rising and mitigation strategies will involve transitions from fossil fuels to bio-energy extraction from biomass with carbon capture and storage (Figure 1).



	Scenario 1: Reduction in energy demand thanks to low-carbon technological innovations.	Scenario 2: Energy demand constantly rising. Shift to bioenergy with carbon capture and storage.
CO_2 emission change in 2030 (% rel to 2010)	-58	4
↳ in 2050 (% rel to 2010)	-93	-97
*Final energy demand** in 2030 (% rel to 2010)*	-15	39
↳ in 2050 (% rel to 2010)	-32	44
from biomass in 2030 (% rel to 2010)	-11	-1
↳ in 2050 (% rel to 2010)	-16	418
Cumulative CCS until 2100 ($GtCO_2$)	0	1218
↳ of which BECCS ($GtCO_2$)	0	1191

Figure 1: Strategies to achieve the net emission reductions that would be required to limit global warming to 1,5°c by 2050. The net amount of emitted carbon dioxide varies between the strategies, as do the relative contributions of bioenergy with carbon capture and storage (BECCS) and the agriculture, forestry and other land use sector. Adapted from the IPCC Special Report from Hoegh-Guldberg et al., 2018.

I.1.2. Soil is a reservoir for CO_2

It is estimated that the carbon contained in the atmosphere increases by 4,3 billion tons each year. Meanwhile, 1 500 billion tons of carbon are stored in soils. Consequently, an annual growth rate of 0.4% in the soil carbon stocks, or 4‰ per year, in the first 30-40 cm of soil, would leverage the atmospheric CO_2 concentration. In line with this observation, the international initiative "4 per 1000", was launched in 2015 at the COP 21 by France, to federate all voluntary stakeholders of the public and private sectors (national governments, local and regional governments, companies, trade organizations, NGOs, research facilities, etc.) to demonstrate that agricultural soils can play a role in the mitigation of climate change. In 2019, INRAE provided a report that highlighted the importance of grasslands, wetlands and forests for carbon storage in soils. In France for example, forest soils store 38 % of the total soil C (Pellerin *et al.*, 2019). Globally, it is estimated that 0,9 Pg C per year are stored in forest soils and dead

wood (Solomon *et al.*, 2007; Venugopal *et al.*, 2016). Soil microbes associated to the rhizosphere, litter and deadwood are crucial actors in the dynamics of C sequestration in soils, as they may promote plant growth, and thereby increase CO_2 assimilation via photosynthesis, or conversely prompt the decomposition of dead organic matter, and thereby release organic compounds that can be assimilated by other organisms and contribute to CO_2 emission via respiration (Lladó *et al.*, 2017). Among soil microbes, wood decay basidiomycetes decompose each year 120 ton/km^2 of wood, resulting in CO_2 efflux (O'Leary *et al.*, 2019). The dynamics of deadwood decay is related to successions of fungal communities and to ligninolytic and cellulolytic activities of wood decayers (Rajala *et al.*, 2012; Ottosson *et al.*, 2014; Valentín *et al.*, 2014; Mäkipää *et al.*, 2017). The dynamics of fungal communities on decaying dead organic matter results for a large part from competition for the resource and the ability of individual species to use the different components of lignocellulosic materials as a carbon source (Hiscox *et al.*, 2018). In particular, white-rot fungi are wood decayers with the capacity to mineralize lignin with ultimate formation of CO_2 and H_2O (Cowling E.B., 1961; Blanchette, 1991). The pivotal role white-rot fungi play in forest ecosystems has stimulated research efforts to understand the enzymatic mechanisms involved in wood degradation.

I.1.3. Plant biomass as a feedstock for renewable carbon: opportunities and drawbacks

Another challenge for sustainability over time is the capabilities of the physical environment to absorb the wastes of human activity. Most commodity chemicals are currently made from platform molecules obtained from fossil fuels. Those platform molecules are produced to supply diverse industrial needs in plastics, synthetic fibers, synthetic rubber, dyes, pigments, paints, coatings, fertilizers, agricultural chemicals, pesticides, cosmetics, soaps, detergents or pharmaceuticals. It is now essential to "*improve the efficiency with which natural resources are used to meet human needs for chemical products and services*" (OECD, 2002) and new technologies must be developed that use renewable resources and avoid the use of persistent, bioaccumulative, toxic, and otherwise hazardous materials.

In this context, the European Commission defends the importance of the bioeconomy as an economy that uses "renewable biological resources like crops, forestry, microorganisms, algae, animals and the residues they generate in order to convert them into value added products, such as food, feed, bio-based products and bioenergy. Thereby, the bioeconomy can contribute to build a more sustainable, innovative, and competitive society while ensuring environmental protection (European Commission, 2012) (Figure 2).

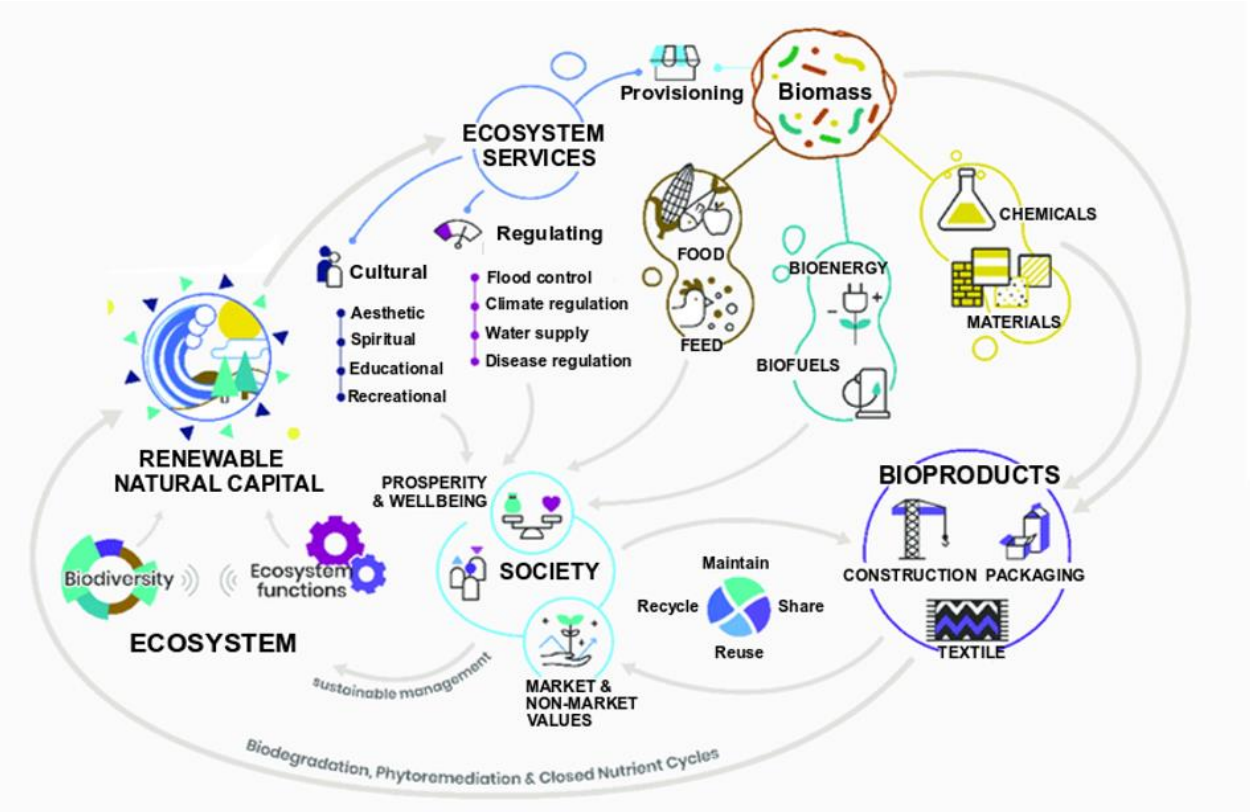


Figure 2: Illustration of circular bioeconomy from Martínez de Arano *et al.*, 2018.

Biomass, that represents organic matters coming from animal, vegetal, bacterial or fungal origin, is gaining lots of interest as an alternative source of carbon. Plant biomass is considered a "clean" resource since roughly, the same amount of carbon is emitted during its conversion to bioenergy and absorbed during the photosynthesis process, thus, preserving the carbon balance. Further reductions in CO_2 emissions can be reached if the carbon generated during the conversion is captured and stored. Bioenergy with carbon capture and storage (BECCS) has the potential for several Gigatons of negative emissions each year (Hoegh-Guldberg *et al.*, 2018). However, the requirement for large biomass supplies, which entails the usage of land surfaces, water and fertilizers is a major concern to the deployment of BECCS (Fridahl & Lehtveer, 2018).

Although some estimations claim that "*There is enough land on the globe to feed 9 billion people and to produce more biomass for energy and material use"* (WBA, 2016), the protection of available lands and forests against desertification, urbanization, and deforestation shapes the potential of biomass for the production of energy and organic chemicals. For example, palm oil production in Indonesia and Malaysia have led to extensive problems of deforestation, loss of wildlife diversity and pollution. Thus, one challenge of bioeconomy is to prevent profit from taking over the natural space. Also, the use of plant biomass should not compete with land

usage for agriculture or food and feed production. One approach is the use of nonedible residues and wastes from forestry, agriculture or agroindustry. Another challenge for sustainable biorefineries is the use of local biomass feedstocks to suppress import transport cost in energy. The combination of those criteria implies that the remaining biomass feedstock is highly diverse and recalcitrant to degradation (e.g. straws, wood sawdust, seed oil press cakes…).

Wood is already contributing significantly to fulfill world's energy demand. In 2019, 85% of the biomass used for energy production were forestry products burnt for heat production (WBA, 2019). In 2017, six pioneer facilities approaching commercial scales (>10 million gallons per year) were producing ethanol from lignocellulosic agricultural residues in USA, Brazil and Italy, which used corn stover, straw, grass or bagasse as a feedstock (Lynd *et al.*, 2017). Although the cost for ethanol production still exceeds the market price of fossil fuel, these pioneer plants allow an evaluation of costs and operability based on experience rather than projection. In France, the Futurol Project (coordinated by J. Tayeb, 2019, https://www.inrae.fr/en/news/2g-ethanol-futurol-technology-almost-market) provided a technology adapted to diverse raw materials (agricultural and forest residues, green urban wastes and whole plants such as miscanthus), allowing the tailoring of the chain according to the geographical and seasonal resources available in the vicinity of the plant.

Although the production of biofuels and bioenergy in a cost-effective way is still a challenge, the co-production of chemicals, materials, food and feed can in principle generate the necessary added value to solve the economic challenges. Several bio-based building blocks (chemicals and polymers) and materials (fibres, starch derivatives, etc.) can be produced from plant biomass through biological, chemical or thermochemical routes (Table 1).

Table 1 : UK Top Bio-based Chemicals in 2017. From (E4tech (UK) Ltd for LBNet, 2017).

Lactic acid	Used to make PLA (polylactic acid), which can be used for biodegradable plastics
2,5-Furandicarboxylic acid (FDCA)	An alternative to PET (polyethylene terephthalate), which is used to make plastic bottles, food packaging and carpets
Levoglucosenone	A safer alternative to toxic solvents used in pharmaceutical manufacturing, flavours and fragrances
5 Hydroxymethyl furfural (HMF)	A building block for plastics and polyesters
Muconic acid	Its derivatives could replace non-sustainable chemicals used in the production of plastics and nylon fibres
Itaconic acid	A replacement for petroleum-based acrylic acid, used to make absorbent materials for nappies and sanitary products, as well as resins used in high-performance marine and automotive components
1,3-Butanediol	A building block for high-value products including pheromones, fragrances, insecticides, antibiotics and synthetic rubber
Glucaric acid	Prevents deposits of limescale and dirt on fabric or dishes, providing a green replacement for phosphate-based detergents
Levulinic acid	Used in the production of environmentally friendly herbicides, flavour and fragrance ingredients, skin creams and degreasers
n-Butanol	Used in a wide range of polymers and plastics, as a solvent in a wide variety of chemical and textile processes and as a paint thinner

Plant biomass consists primarily of carbohydrate polymers, cellulose and hemicellulose, that can be degraded to generate fermentable monosaccharides which are further converted by yeasts into bioethanol (Figure 3). However, the conversion of lignocellulosic materials is facing the complexity of the lignocellulosic structure. In particular, the highly recalcitrant lignin polymer and the crystalline regions of cellulose microfibrils limit the breakdown process and make it energy consuming (Srivastava *et al.*, 2015). Thus, in a bio-based economy, efforts are made to find environment-friendly and economically sustainable methods for biomass pre-treatment. One solution was found in many microorganisms that have evolved abilities to efficiently break down all plant cell wall (PCW) polymers and access carbon sources. Among those microorganisms, fungi became the predominant source of enzymes currently used on an industrial scale for biomass transformation (van den Brink & de Vries, 2011). For example, *Trichoderma reesei*, a fungal strain that was first isolated on American soldiers' uniforms during World War II in the Solomon Islands, is a prolific producer of cellulases and hemicellulases, the

key components of industrial cocktails dedicated to biomass degradation. This species has undergone intensive strain improvements to maximize enzyme production yields and optimize the produced enzyme cocktails, notably by upgrading, through synthetic biology, with enzymes from other biomass-degrading fungi (Seidl & Seiboth, 2010; Filiatrault-Chastel *et al.*, 2019).

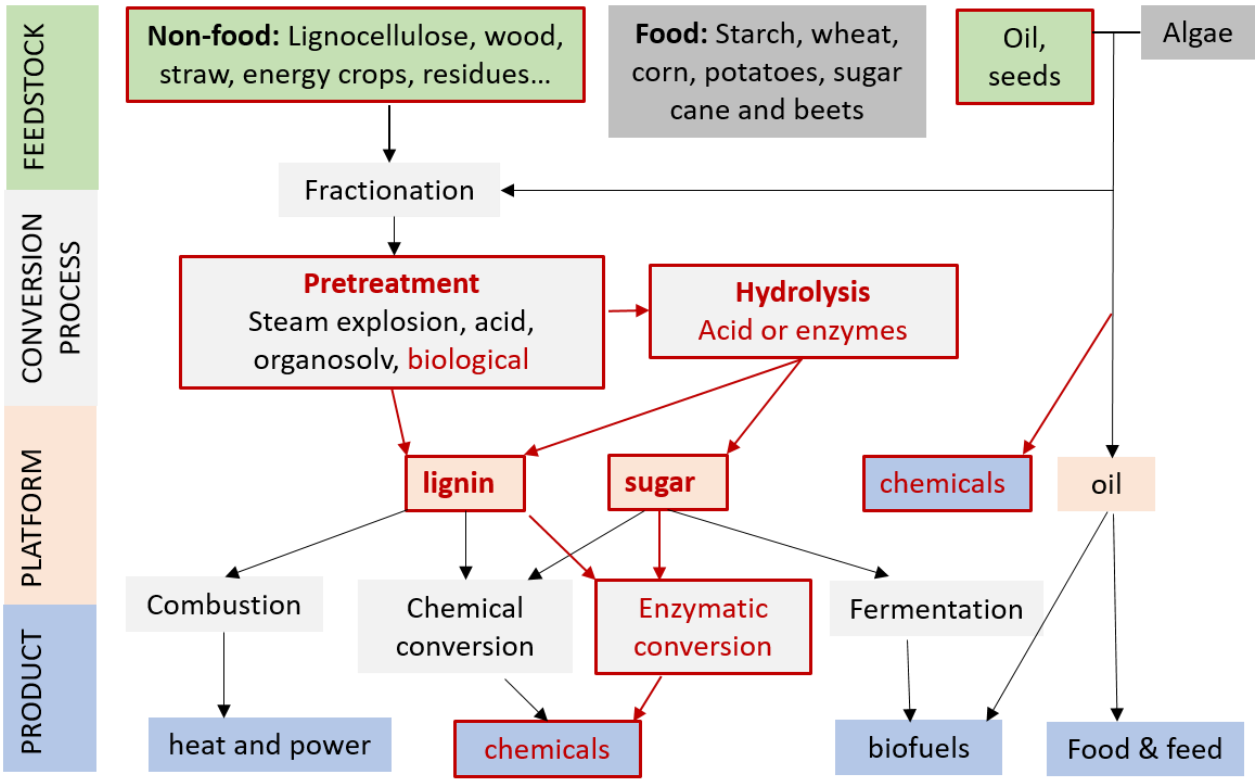


Figure 3: Technological steps for biomass conversion. The red boxes highlight the processes in which fungi or fungal enzymes can be used. Adapted from Stichnothe *et al.*, 2020.

I.1.4. Filamentous fungi as enzyme factories

While several species have been identified as efficient factories to produce enzymes of interest for biotechnology, less efficient species can undergo strain improvement using synthetic biology, thus shaping their potential to produce a desired product (Gupta *et al.*, 2016; Park *et al.*, 2017). For example, model strains for biofuel technologies can be constructed to produce high yields of biomass-degrading enzymes, in an inexpensive growth substrate and with a minimal production of undesirable by-products (Teotia *et al.*, 2016). Strain improvement can be achieved via genetic modification by gene additions or via increased expression of the enzymes of interest (Nayak *et al.*, 2006). For example, in *Pleurotus ostreatus*, the overproduction of β-glucanases was reached by replacement of the native promotor with a promotor coming from *Aspergillus nidulans* (Chai *et al.*, 2013). Another example is the deletion

of a PepA protease in *Aspergillus awamori* for large production of the food sweetener thaumatin (Moralejo *et al.*, 2002). Also, Miyauchi *et al.*, 2014 generated a high xylanase-producing strain of *Trichoderma reesei* by introducing multi-copies of the bacterial xynB gene controlled by different native promoters. Many other genome-editing techniques have been rapidly developed for filamentous fungi and efforts are underway to improve biomass degradation by targeting all PCW components via a single fungal strain.

To conclude, altogether, recent findings and ongoing developments highlight fungi as pivotal actors in the mitigation of CO_2 emissions, both in natural environments and as toolkits for biotechnologies centered on renewable carbon.

Chapter II INTRODUCTION

II.1. Plant cell wall degrading enzymes

The plant cell wall (PCW) consists mainly of a network of cross-linked polysaccharide polymers and polyphenols known as lignocellulose (cellulose, hemicellulose and lignin) (Figure 4). Those building blocks confer to the plant its rigidity, resistance against degradation and protection from microbial attack (Higuchi, 1990). In addition to lignocellulose, the PCW is formed of minor fractions of pectin, lipids, glycoproteins and minerals.

Cellulose, which is the main component of PCW (35-50%) is considered as the most abundant biopolymer on earth. It consists of several hundred to over ten thousand linear chains of D-glucopyranose units which are held together by β-1,4-glycosidic bonds. Those basic cellulose chains are stabilized together via an extensive network of inter- and intramolecular hydrogen bonds, which form the microfibrils. Cellulose microfibrils contain regions with organized crystalline cellulose and less organized amorphous regions (Ding & Himmel, 2006). The crystalline cellulose is strengthened by strong interchain hydrogen bonds and by a protective layer of hemicellulose that make it highly resistant to enzymatic hydrolysis and thus recalcitrant to degradation (Beckham *et al.*, 2011; Couturier *et al.*, 2018). The ratio of crystalline to amorphous cellulose can vary between the plant species.

Hemicellulose represents about 20 to 35% of lignocellulosic biomass and encompasses heteroxylans, heteromannans, xyloglucans and mixed-linkage glucans. The main chains are branched with short lateral chains consisting of L-arabinose, 4-O-methyl-glucuronic, D–galacturonic, or D-glucuronic acids. In PCW, hemicelluloses may interact with lignin through covalent ester bonds, which insures a coating protection of the cellulose microfibrils against biodegradation. The structure and composition of hemicellulose can vary among different cells of an individual plant or between different plant types (Scheller & Ulvskov, 2010). For example, the majority of hemicelluloses from softwood (from gymnosperm trees) are glucomannans, whereas hemicelluloses from hardwood (from angiosperm trees) consist mainly of xylans (Himmel *et al.*, 1994).

Lignin is the second most abundant source of carbon on earth after cellulose. Lignin is a complex polymer synthetized mainly from three aromatic alcohols: *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol, which are incorporated into the polymer as *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units, respectively (Higuchi T., 1998; Boerjan *et al.*, 2003; Figure 4). The lignin composition varies between softwood and hardwood; softwood lignins consist almost exclusively of G-type lignin, whereas hardwood lignins consist of S and G units (H units are minor components in both) (Campbell & Sederoff, 1996).

Overall, the ratio, structure and chemical composition of cellulose, hemicellulose and lignin differ strongly between plant species, tissues and the growth phases of the plant. For example, softwood (e.g., pine and spruce wood) has higher lignin content than hardwood (e.g., birch, aspen, and oak wood), whereas the amount of cellulose in softwood is smaller than that in hardwood (Rytioja *et al.*, 2014). Also, agricultural straws have higher content in hemicellulose as compared to wood (Olsson & Salmén, 2004).

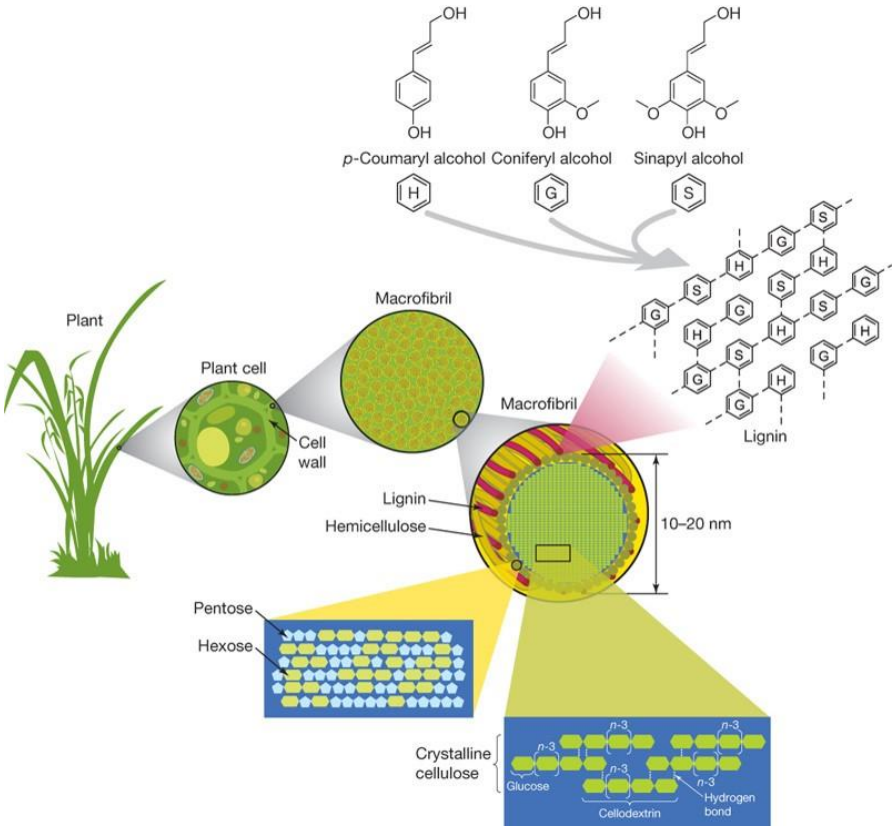


Figure 4: Lignocellulose composition (from E.M. Rubin, 2008).

As a consequence of the huge chemical complexity that can be found in PCW polymers, the organisms that use plant tissues as a carbon source have adapted a large diversity of PCW degrading or modifying enzymes (PCWDE). Those enzymes are classified, together with all carbohydrate-active enzymes (CAZymes) identified to date, in the CAZy database (http://www.cazy.org/; Lombard *et al.*, 2014), into families according to their amino acid sequence and structural similarity. More than 300 protein families are currently described in CAZy, divided into 6 major classes. The largest class is represented by glycoside hydrolases (GHs), which hydrolyse the glycosidic bond between carbohydrates (Henrissat & Davies, 1997). The other classes are: the glycosyl transferases (GTs), which form glycosidic bonds via the transfer of a carbohydrate from a donor molecule to an acceptor molecule, carbohydrate esterases (CEs), which hydrolyse ester bounds in hemicelluloses, cutin and suberin,

polysaccharide lyases (PLs), which cleave uronic acid-containing polysaccharide chains (Cantarel *et al.*, 2009), auxiliary activity enzymes (AA), which are oxidoreductases that contribute directly or indirectly to the degradation of polysaccharides and lignin, and carbohydrate binding modules (CBMs), which are non-catalytic modules with the capacity to bind to soluble and crystalline carbohydrates. A wealth of enzymes that cleave PCW polymers have been identified in the fungal kingdom (Figure 5; Table 2). Yet, the landscape of fungal PCWDE remains largely untapped. Some of the enzymes classified in the CAZy database still call for functional characterization, and new CAZymes are continuously being discovered (e.g. Couturier *et al.*, 2018; Filiatrault-Chastel *et al.*, 2019).

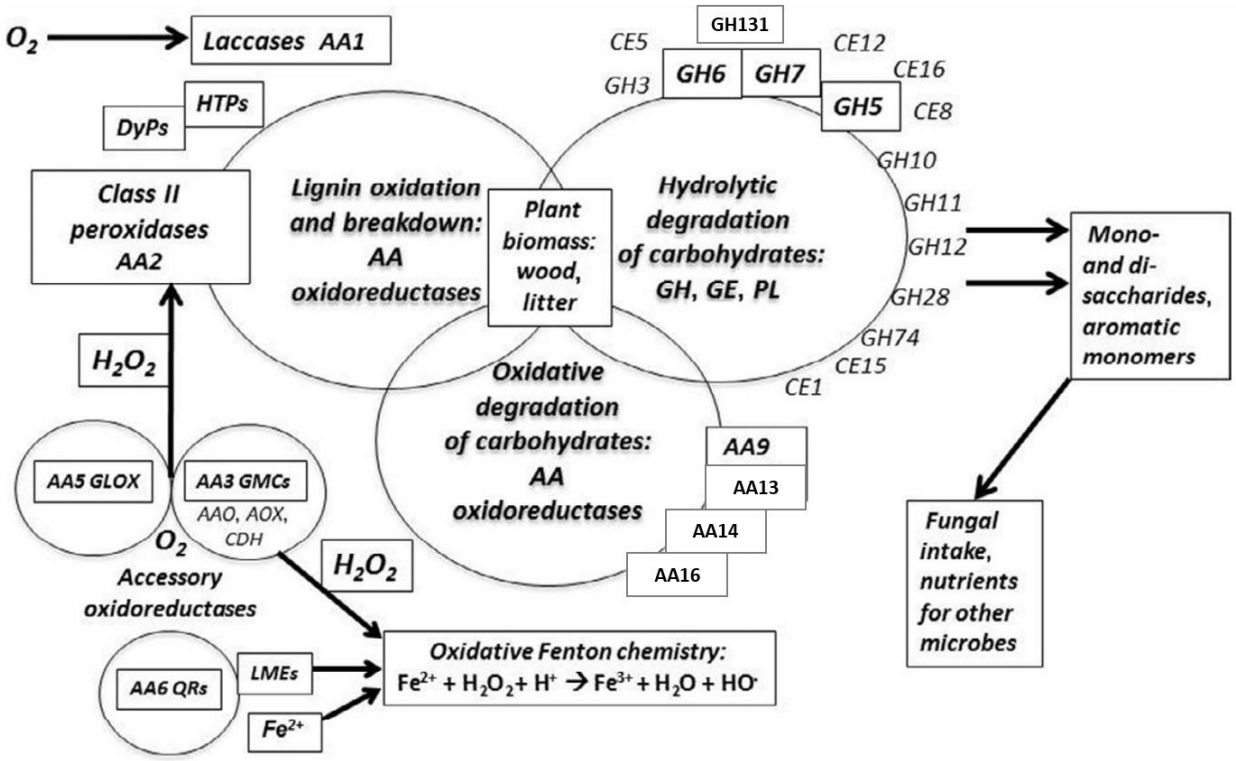


Figure 5: Fungal secreted CAZymes involved in the degradation of plant cell walls. GH: glycoside hydrolase; CE: carbohydrate esterase; AA: auxiliary activity enzyme; LME : lignin-modifying enzyme ; QR : quinone reductase ; GMC : glucose-methanol-choline oxidoreductase; AAO: aryl alcohol oxidase; AOX: alcohol oxidase; CDH: cellobiose dehydrogenase; DyP: dye-decolorizing peroxidase; HTP: heme-thiolate peroxidase. (Adapted from Lundell et al., 2014).

Table 2: Fungal CAZymes active on plant cell walls. The activities were reported in (Henriksson et al., 2000; Cantarel et al., 2009; Gruber & Seidl-Seiboth, 2012; Floudas et al., 2015; Sipos et al., 2017; Couturier et al., 2018; Andlar et al., 2018; Filiatrault-Chastel et al., 2019)

Substrates	Enzyme classes	Enzymatic activity	Cazy family
Lignin	copper radical oxidase	Laccase	AA1_1
	Class II heme peroxidase	Manganese peroxidase	AA2
		Lignin peroxidase	
		Versatile peroxidase	
Hemicellulose	Glycoside Hydrolases	Endo-1,4-β-xylanase	GH5_22, GH8, GH10, GH11, GH30_7
		xyloglucanase	GH74, GH44
		Endo-β-1,4-mannanase	GH5_1, GH5_7, GH26, GH113, GH134
		Endo-α-1,5-arabinanase	GH93
		β-1,3-glucanase	GH16
		α-L-arabinofuranosidase	GH43, GH51, GH54, GH62
		β-Glucuronidase	GH115, GH2
		α-1,2-glucuronidase	GH67, GH115
		β-mannosidase	GH2, GH5_2
		β-Galactosidase	GH35, GH53
		α-Galactosidase	GH27, GH36
		β-xylosidase	GH52, GH54, GH120, GH30_1, GH39
		α-L-fucosidase	GH29, GH95, GH141
	Carbohydrate Esterases	Acetylxylan esterase	CE1, CE2, CE3, CE4, CE6, CE16
		Cutinase	CE5
		Glucuronyl methyl esterase	CE15
	Auxiliary activity	Lytic polysaccharide monooxygenase	AA14
Cellulose	Glycoside Hydrolases	Endoglucanase	GH5,4, GH5_5, GH12, GH45, GH74, GH131
		cellobiohydrolase	GH6, GH7, GH5_1, GH48
		β-Glucosidase	GH1, GH3, GH30_1, GH5_7, GH5_22
	Auxiliary activity	Lytic polysaccharide monooxygenase	AA9, AA16
Pectin	Glycoside Hydrolases	Polygalacturonases	GH28, GH78
		β-glucuronyl hydrolase	GH88, GH105
		α-L-rhamnosidase	GH78, GH106
		β-1,4-galactanase	GH53
	Polysaccharide Lyases	Polygalacturonate lyases	PL1, PL3, PL9
		Rhamnogalacturonan lyase	PL4, PL11
	Carbohydrate Esterases	Rhamnogalacturonan acetylesterase	CE12
		Pectin methylesterase	CE8
Cutin	Carbohydrate Esterases	Cutinase	CE5
oxido-reductases	Auxiliary activity	FAD-dependent (GMC) oxidoreductases	AA3
		vanillin alcohol oxidase	AA4
		copper radical oxidase	AA5
		benzoquinon reductase	AA6
		glycooligosaccharide oxidase	AA7
		pyrroloquinoline quinone-dependent oxidoreductase	AA12
		Cellobiose dehydrogenases	AA8-AA3
carbohydrate binding modules		Xylan, galactan	CBM13
		cellulose	CBM1, CBM63
		pectin	CBM67

II.2. Fungal biodiversity as a reservoir of plant cell wall degrading enzymes

II.2.1. Fungal biodiversity as a result of their adaptation to diverse habitats

The oldest fossil filamentous fungus on record dates back to one billion years ago (Blackwell M., 2011) and to date, as many as 12 million fungal species are estimated to exist on earth. The apparition of this enormous number of species reflects the long evolution they went through. They evolved over time different strategies to survive in a wide range of ecological niches with extreme physicochemical parameters such as temperature (from -40°C to +75°C), pH, water availability, gaz concentration… Specialized fungi have been discovered in extreme cold, dry, acidic, salty, and deep-sea habitats, while others can only proliferate in environments with moderate physicochemical properties. Studies have shown that those latter are much more likely to diversify than specialist species that evolved constrained adaptation to local conditions (Buckling *et al.*, 2003; Gostinčar *et al.*, 2009).

Fungi are heterotrophs that use extracellular digestion to degrade the organic matter they use as a source of carbon and energy. The efficiency of fungal survival is partly related to the enzymatic machineries they use to access carbon, which vary widely based on the composition of the organic matter of the ecological niches they evolved on and the strategies they adapted during evolution.

In the past few years, lots of research efforts have been focusing on the discovery of new enzymes that degrade PCW polysaccharides into metabolizable sugars. The discovery of such enzymes can be elucidated through the comparative analysis of lignocellulose-degrading fungi having different substrate preferences, different ecologies, or belonging to different taxonomical groups. Fortunately, in recent years, fungal genomes have been sequenced, assembled and annotated in an amazing speed reaching an unprecedented scale (Kuo *et al.*, 2014). Indeed, the beginning of fungal genomics era happened in 1996 after the sequencing of the first eukaryotic genome, from *Saccharomyces cerevisiae*, and to date, NCBI records 6993 fungal genomes publicly available. One of the latest projects in fungal genomics, the Project 1000 Fungal Genomes, aims to comprehensively document fungal genomic diversity by sequencing a large variety of fungi across the fungal tree of life, including fungi from yet unstudied taxonomical groups (Grigoriev *et al.*, 2014). The enormous amount of newly produced genomic data has become easily accessible thanks to integrative databases developed to store, analyze, and compare fungal genomes (e.g. the MycoCosm repository developed by the Joint Genome Institute (https://mycocosm.jgi.doe.gov/), FungiDB developed by the American Bioinformatics Resource Center; https://fungidb.org/ or Ensembl Fungi developed by the European Molecular Biology Laboratory's European Bioinformatics Institute (https://fungi.ensembl.org).

II.2.2. Fungal adaptation to plant hosts and plant substrates

The earliest land plants, as well as their closest relatives, the streptophyte algae contained cellulose and pectin in their cell walls and lacked lignin (Popper *et al.*, 2011). Recent genome sequencing efforts have shown that early diverging fungi did possess pectinase and cellulase genes in their genomes (Lange *et al.*, 2019), suggesting that pectin and cellulose degrading enzymes have been part of the ancestral fungal toolkit for breaking down plant material and that fungus-plant associations have originated around 750 million years ago, the approximate time of the emergence of pectin-containing streptophytes (Douzery *et al.*, 2004; Zimmer *et al.*, 2007; Parfrey *et al.*, 2011) (Figure 6). Later, pectinase genes have undergone rapid duplications in organisms that adopted a plant-based nutrition (Chang *et al.*, 2015). In a recent study, Anasontzis *et al.*, (2019) analyzed the distribution and enzymatic activity of GH131, a CAZy family only found in fungi and plant parasitic oomycetes. Several features suggested that GH131s could also result from fungal adaptations to land plants. GH131s have endo-β-1,3 and endo-β-1,4 activity, suggesting they play a role in PCW breakdown or modification by cleaving the mixed hemicellulose β-1,3/1,4 bonds or the cellulose polymers. GH131s probably appeared in a common ancestor of Dikarya, a subkingdom of fungi that arose around 662 Mya (Floudas *et al.*, 2012), shortly after land plants [700 Mya (Heckman *et al.*, 2001)] and GH131 gene copy numbers are higher in the genomes of fungi that colonize plant tissues (saprotrophs, pathogens and symbionts). Some of the copies have the catalytic domain fused to a CBM1, a carbohydrate binding domain that has an affinity to crystalline cellulose and directs CBM1-associated enzymes to potentiate cellulolytic activities on insoluble substrates (Lehtiö *et al.*, 2003; Fong *et al.*, 2016). During plant tissue colonization, GH131 genes were found to be co-expressed with genes coding for other PCW-degrading enzymes. Finally, the co-secretion of GH131 with AA9 Lytic Polysaccharide Monooxygenases (LPMOs), cellobiohydrolases and other endoglucanases (Miyauchi *et al.*, 2016) supported the fact that GH131 act in synergy with other enzymes, for PCW modification and contributed to the adaptation of fungi to growth in plant tissues.

Another example of fungal adaptation to land plants is the evolution of class II peroxidases (PODs) genes, which are specific to the Agaricomycetes, a class of fungi among Basidiomycota. PODs have been classified based on the presence of key catalytic residues into four major groups, including three ligninolytic forms— manganese peroxidase (MnP), versatile peroxidase (VP) and lignin peroxidase (LiP), — and a fourth POD type, generic peroxidase (GP), which include non-ligninolytic low–redox potential peroxidases. Using phylogenomics, Floudas *et al.*, (2012) showed that the common ancestor of PODs was a non-ligninolytic GP that acquired the Mn(II)-binding site specific of MnPs, formed by two glutamates and one aspartate. This allows MnPs to catalyze the oxidation of phenolic lignin and lignin compounds via Mn3+ chelates (Gold *et al.*, 2000). They proposed that this evolution of non-ligninolytic GP to ligninolytic MnP happened during the late Carboniferous period, after the accumulation of lignin-containing

biomass on earth (Floudas *et al.*, 2012). In some lineages, the catalytic properties of MnP further evolved in parallel with the evolution of lignin complexity in plants. The lignin of angiosperm trees is composed of G and S phenylpropane units and is chemically more complex than the lignin of gymnosperm trees. Recent studies on the resurrection of ancestral Polyporales enzymes showed that the modification of lignin composition in angiosperms drove a fascinating coevolution of fungal MnP to VP, in which a surface tryptophan confers the ability to abstract electrons from nonphenolic lignin and transfer them to the cofactor via a long-range electron transfer (LRET) pathway (Ayuso-Fernández *et al.*, 2018). Finally, LiPs appeared in the Polyporales and Agaricales orders by convergent evolution from different VP ancestors after the loss of the Mn(II) binding site (Ayuso-Fernández *et al.*, 2018) (Figure 6). In a recent phylogenomics study on 33 Agaricales genomes, Ruiz-Dueñas *et al.* (2020) further unraveled the adaptation of Agaricales to diverse ecological niches (buried wood, decayed wood, leaf litter and grass litter), that paralleled the appearance of MnPs and VPs with diverse Mn(II) binding sites.

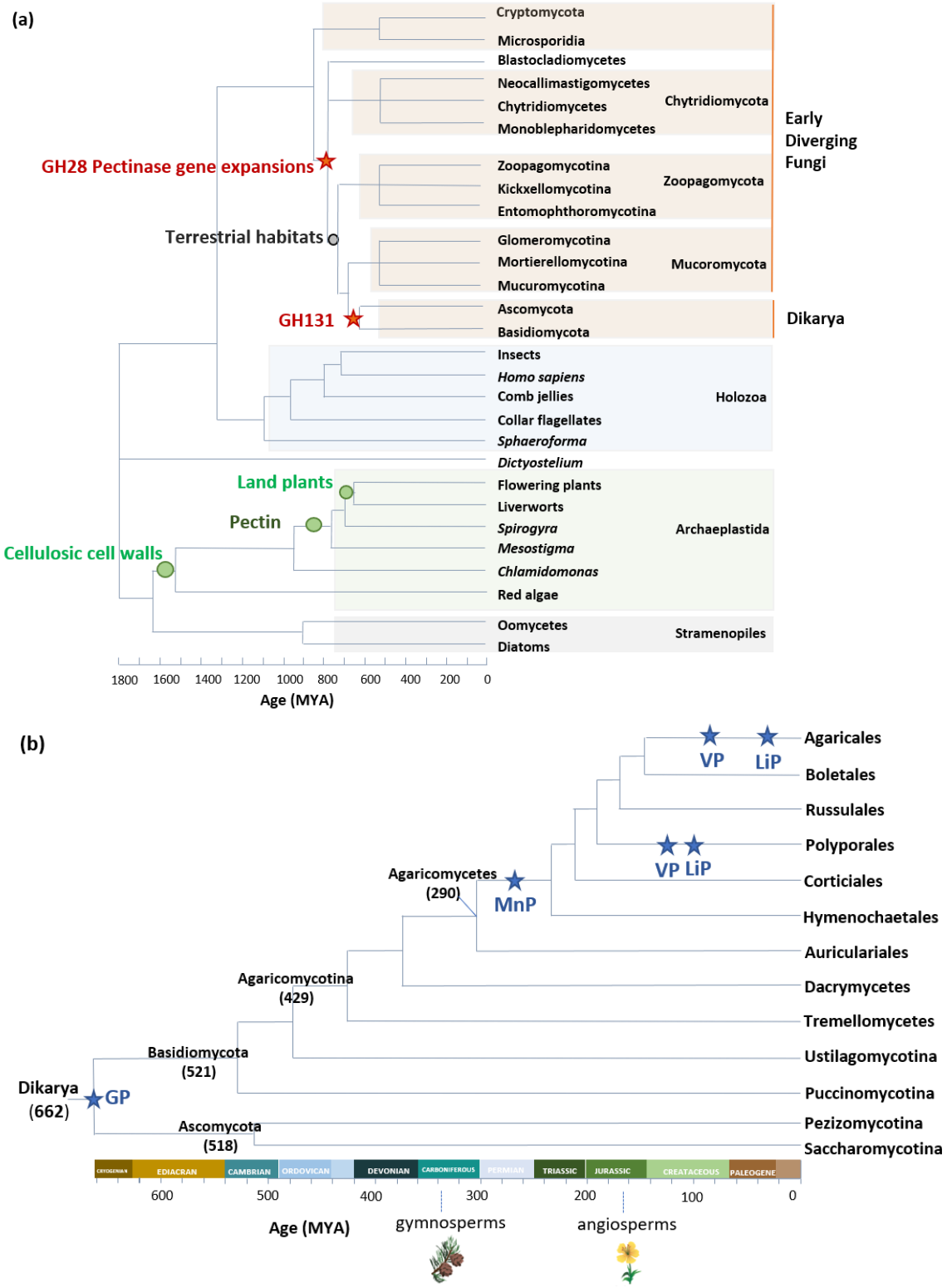


Figure 6 Early evolutionary events for (a) GH28 and GH131 enzymes, and (b) Class II peroxidases. The estimated ages of fungal taxa are from Floudas et al., 2012; Chang et al., 2015; Berbee et al., 2017; Spatafora et al., 2017; Ayuso-Fernández et al., 2018; Ruiz-Dueñas et al., 2020).

II.2.3. Fungal adaptation to wood decay

Based on their strategies to acquire saccharide nutrients, fungi living on plant tissues can be classified into three major ecological groups: 1) pathogens withdraw the saccharides from living organisms and eventually cause the death of the infested tissues, 2) symbiotic fungi sustain a beneficial association with their host plants. They can be endophytes, for which the exchanged benefits are still unclear, or mycorrhizal and play a role in facilitating the plant uptake of water and minerals such as nitrogen and phosphorus, or 3) saprotrophic and take their nutrients from the decomposing organic matter coming from dead organisms. Although fungal ecologies mostly follow one of those three categories, some fungi were found to have an intermediate, transitory or dual lifestyle suggesting the presence of a continuum in fungal lifestyles rather than a clear-cut classification (Sánchez-García *et al.*, 2020). For example, some species were found to be able to grow as pathogens and saprobes, like the tree associated fungi *Armillaria* spp. or *Heterobasidion annosum* (Anderson & Stasovski, 1992; Olson *et al.*, 2012). In addition, the endophytic lifestyle is now described as a transitory lifestyle where endophytic species are potentially latent pathogens (Carroll, 1988; Begoude *et al.*, 2011; Andrew *et al.*, 2012). Those fungal abilities to switch between different lifestyles are made possible by the wide enzymatic toolkits that their genomes offer. Therefore, the larger the number of fungi studied, the more varieties of enzymatic systems for PCW degradation or modification can be discovered.

Saprotrophic wood decay fungi

Saprotrophic fungi feed on the carbon retrieved from dead tissues. Since two-third of the organic carbon on earth is fixed in non-living biomass, those saprotrophic fungi are major contributors to the carbon cycle, and are estimated to play a role in the release of 85 billion tons of carbon each year (Gilbertson R.L., 1980).

Among saprobes, wood-decay fungi can be classified into three major types based on the aspect of the wood following fungal colonization: white-rot, brown-rot or soft-rot fungi. White rotters, the dominant wood-degrading species, hold their name to the white and fibrous decayed wood they produce (Figure 7). They are rich in PCWDE and therefore have the potential to degrade the entirety of the PCW polymers, including cellulose, hemicellulose, pectin and lignin. In contrast, the wood decayed by brown-rot fungi shows dry and brown cubic debris in which the cellulosic and hemicellulosic compounds are degraded but lignin is oxidized and only partially degraded. These fungi have reduced portfolios for PCWDE and produce reactive oxygen hydroxyl radicals via the Fenton reaction to cleave the PCW polymers (Blanchette R.A., 1984; Martínez *et al.*, 2005; Goodell *et al.*, 2008). Soft-rot fungi typically attack wood with reduced lignin content in high moisture conditions and can cause greyish discoloration and fragmentation of the decayed wood. Although this classification facilitates the comparative

analyses of wood decay types, it is noteworthy that the white-rot/brown-rot dichotomy tends not to be representative of the diversity of fungal degradation strategies (Schilling *et al.*, 2020). Indeed, the presence of ligninolytic POD genes in a genome was first correlated to the classification of the species as white-rotter (Floudas *et al.*, 2012; Ruiz-Dueñas *et al.*, 2013; Nagy *et al.*, 2016). However, some species do produce a white-rot phenotype although no ligninolytic PODs were identified in their genome (Riley *et al.*, 2014; Floudas *et al.*, 2015). One explanation for this discrepancy is that the recorded observations in natural environments or in laboratory conditions might not cover the diversity of wood decay phenotypes associated to a species. Alternatively, it has been suggested that those fungi are in transition between white-rot and brown-rot decay type, similar to the transitions that took place independently at least three times during Agaricomycetes evolution (Kohler *et al.*, 2015). Besides, each rot type is known to deploy a characteristic panel of enzymatic and chemical mechanisms for the degradation of PCW polymers (reviewed in Lundell *et al.*, 2014). Brown-rot fungi possess a significantly smaller set of PCWDE as compared to white-rot fungi (Figure 8). Notably, CAZymes related to the depolymerization of crystalline cellulose (Lytic Polysaccharide Monooxygenases LPMO from CAZy family AA9) or lignin (Class II peroxidases POD, multicopper oxidases MCO and H_2O_2-generating GMC-oxidoreductases) are expanded in white-rot fungi as compared to brown-rot fungi (Floudas *et al.*, 2012; Ruiz-Dueñas *et al.*, 2013). In addition, brown-rot fungi show different gene expression patterns, and the induction of ferric reductase and heme-thiolate peroxidase/peroxygenase genes in the early stages of wood degradation (Zhang *et al.*, 2019).

Besides enzymatic toolkits for PCW degradation, wood decay fungi have also developed intracellular detoxification systems involved in the elimination of xenobiotics and other potentially toxic compounds released or generated during the degradation of wood. Of note, the enzymes involved in detoxification are of great interest in medicine, biotechnology and bioremediation. For example, cytochrome P450 have high potential in the production of drugs, drug intermediates, dyes or pesticides (Kumar S., 2010), and glutathione transferases can be interesting for herbicide and heavy metal detoxification or hormone biosynthesis (Kumar & Trivedi, 2018).

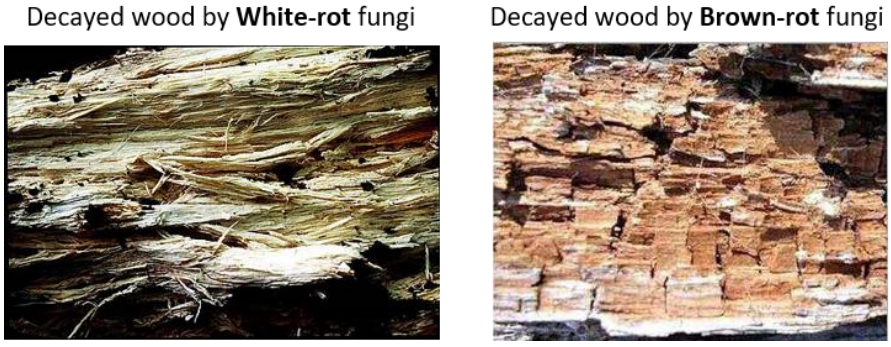


Figure 7: Different aspects of wood decayed by saprotrophic fungi (pictures from Tom Volk, University of Wisconsin (left) and Jim Deacon, The University of Edinburgh (right)).

Litter decay fungi

Among saprotrophic fungi, litter-decomposing species in forests and grasslands have evolved a stronger enzymatic toolkit for PCW degradation as compared to the wood-decayers, and are notably rich in enzymes acting on the lignin and polysaccharide fractions of the PCW. For example, litter-decomposing species from the Agaricales order have significantly higher gene-copy numbers of AA9 LPMO, laccases and PCWDE genes appended to a CBM1 (Figure 8). It is suggested that the CBM-appended catalytic domains could prevent enzyme leaching in the loose environment of litter, as compared with compact wood. In addition, the adaptation of Agaricales to diverse ecological niches (e.g. forest litter, grass litter, dead wood, buried wood), has paralleled a surprisingly high diversification of ligninolytic peroxidases, evolving from ancestral short MnP to long MnP, VP and LiP, with the apparition of several POD types with non-canonical catalytic amino acids (Ruiz-Dueñas *et al.*, 2020).

Mycorrhizal fungi

Mycorrhizal fungi establish a mutualistic association with host plants. They colonize the plant to extract the required carbon for their nutrition, while the plant benefits from the water, phosphorus and nitrogen conveyed from the soil by the fungus (Kosuta *et al.*, 2003). To establish this association, mycorrhizal fungi developed during more than 400 million years of co-evolution with plants, strategies to colonize plant tissues. Mycorrhizal fungi are classified into four main types according to the morphological diversity of the symbiotic structures. Arbuscular mycorrhizae (AM) are formed by Glomeromycetes that develop arbuscules inside the host cells. Ectomycorrhizae (ECM) are formed by higher fungi that produce a labyrinthine hypha between host cells. Orchid mycorrhizae fungi (ORM) form coils of hyphae (pelotons) that penetrate within cells in the plant family Orchidaceae. Finally, ericoid mycorrhizae (ERM) are formed by some ascomycete fungi and characterized by pelotons entering within very thin roots of Ericaceae (Brundrett M., 2004).

ECM fungi have a reduced set of CAZymes as compared to saprotrophic fungi, which allows tuned modifications of PCW for nutrient exchange with minimal disruption (Figure 8). Their enzymatic portfolio is somehow similar to that of brown-rot fungi (Rineau *et al.*, 2012, 2013), both having lost most of the plesiomorphic enzymatic machineries of their white-rot ancestors. For example, GH6 GH6 (cellobiohydrolase) genes, CE1 (acetylxylan esterase) genes related to hemicellulose degradation, pectate lyases and lignin-acitve PODs are absent in ECM species (Kohler *et al.*, 2015; Ruiz-Dueñas *et al.*, 2020). Conversely, GH5_5 (β-1,4-endoglucanases), GH12 (β-1,4-endoglucanases), AA9 LPMOs and GH28 (polygalacturonases), could play a role in the first steps of the colonization process, during the penetration of the plant root cortex (Veneault-Fourrey *et al.*, 2014), as demonstrated by the reduced numbers of mycorrhizae formed by *Laccaria bicolor* strains knocked-down for GH5_5 (Zhang *et al.*, 2018). In contrast to this observed loss of PCW degrading enzymes, mycorrhizal genomes seem to have undergone

expansions of laccase genes that appear to be related to fruiting body development in some species, (Morin *et al.*, 2012).

Recent genome sequencing efforts have shown that AM fungi also have reduced portfolios for genes coding for PCWDE. Conversely, the genomes of ascomycete ERM, ORM and endophyte fungi are rich in PCWDE genes, which is proposed to mirror a versatile lifestyle, and the ability to behave as saprotroph in the absence of a plant host (Perotto *et al.*, 2018).

ECM genomes also show the expansion of genes coding for small secreted proteins (<300 amino acids; SSP) where some are shared with saprotrophic species, while others appear to be ECM-specific SSPs. Although the functions of a large number of SSPs are yet unknown, some have been shown to act as effectors that hijack host defenses or manipulate plant cell signaling. For example, Plett *et al.* (2014) have shown that the protein MiSSP7 (mycorrhiza-induced SSP7) from *L. bicolor* acts in the host-plant nuclei to suppress plant defense responses and allow the development of the Hartig net required for the establishment and functioning of the symbiosis. Other SSPs are involved in the differentiation of fungal structures (such as the fruiting body) or in the competition between fungi for the host (Trejo-Hernández *et al.*, 2014).

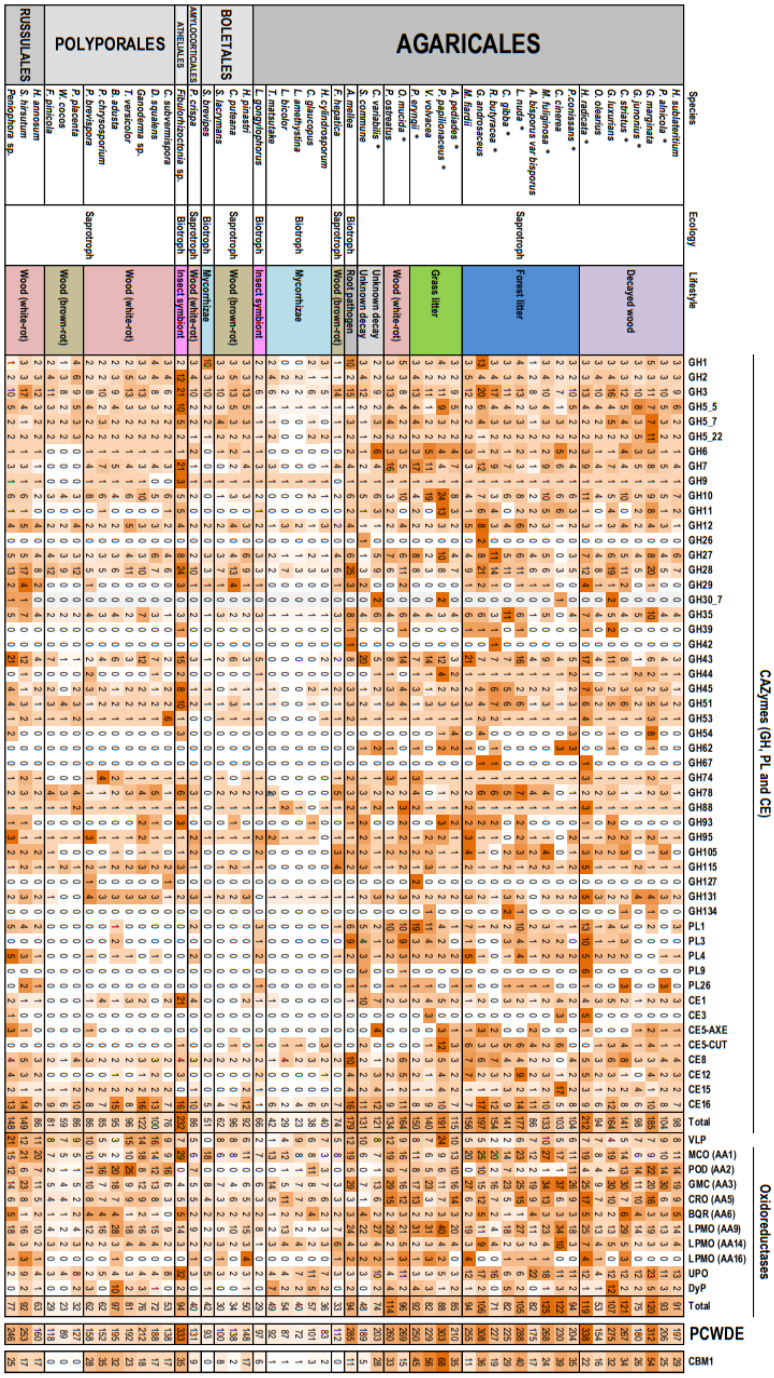


Figure 8: Distribution of CAZymes, oxidoreductases, CBM1 and versatile lipases in the genomes of 52 Agaricomycetes species having diverse lifestyles. The gene counts for 38 glycoside-hydrolase (GH), 5 polysaccharide-lyase (PL) and 7 carbohydrate-esterase (CE) families are shown together with gene counts for multicopper oxidases (MCO; AA1), POD (AA2), GMC-oxidoreductases (AA3), copper radical oxidases (CRO; AA5), benzoquinone reductases (BQR, AA6), LPMO (AA9, AA14 and AA16), unspecific peroxygenases (UPO) and DyP peroxidases. In each column, the color goes from orange, for the species with the highest number of genes, to white when a gene is absent. From Ruiz-Dueñas *et al.*, 2020.

II.3. Polyporales (Basidiomycetes) as a reservoir of PCW modifying enzymes

The order Polyporales, which represents 1.5% of all described fungal species (Kirk *et al.*, 2009), is a taxonomic order in the Agaricomycetes, a highly diverse and large group of mushroom-forming fungi within the phylum Basidiomycota (Figure 6). They are widely distributed in boreal, temperate, tropical and subtropical areas. The phylum Polyporales is estimated to have emerged about 194 mya (Song & Cui, 2017) and has been subdivided phylogenetically into two main lineages which include the core polyporoid, gelatoporia, antrodia, phlebioid, and residual polyporoid clades among others (Justo *et al.*, 2017). Polyporales is one of the most intensively studied clades of fungi with 5482 publications having the keyword "Polyporales" recorded in PubMed to date. Among the famous Polyporales, *Fomes fomentarius,* traditionally used as tinder material and for spiritual purposes, has been found among the equipment of the 5000-year-old mummy of Otzi the Iceman, which was found in an alpine glacier (Figure 9) (Peintner *et al.*, 1998). Otzi also carried with him another Polyporales fungus, *Fomitopsis betulina*, speculated to have been used as a laxative to expel whipworm. These two species have a long history of medicinal use and interest for isolation of secondary metabolites. Another famous Polyporales fungus is *Ganoderma lucidum,* which has been recognized as a medicinal mushroom for over 2000 years and is regarded as the "herb of spiritual potency" in Chinese culture.

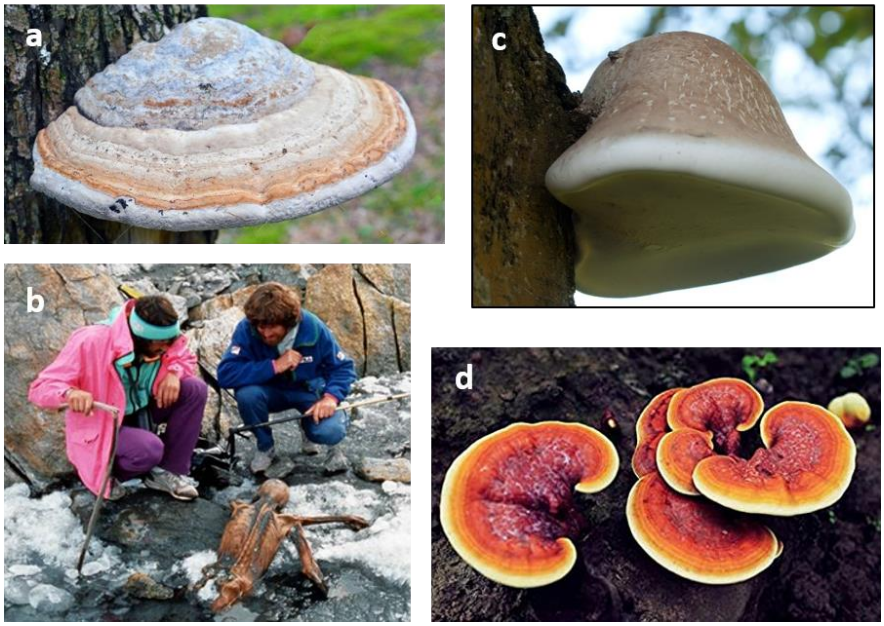


Figure 9 a. *Fomes fomentarius,* **b.** Otzi the iceman found in an alpine glacier **c.** *Fomitopsis betulina* and d. *Ganoderma lucidum.*

Furthermore, Polyporales species are showing a great interest because of their remarkable efficiency in wood degradation and high potential for the isolation of efficient PCWDEs. Notably, recent genome sequencing efforts have helped in the discovery of new families of LPMOs that cleave crystalline cellulose (Filiatrault-Chastel *et al.*, 2019) or xylan polymers coated on cellulose microfibrils (Couturier *et al.*, 2018). In addition, two dehydrogenases, an aryl-alcohol dehydrogenase and a glucose dehydrogenase, were identified from the genome of the white-rot fungus *Pycnoporus cinnabarinus* (Piumi *et al.*, 2014; Mathieu *et al.*, 2016).

The plesiomorphic decay type for Polyporales is white-rot (Floudas *et al.*, 2012). Species from the antrodia lineage and a few genera within Polyporales transitioned to the brown rot type (Justo *et al.*, 2017). Within white-rot Polyporales, different species can degrade lignin with different degrees of cellulose preservation. This diversity in degradation abilities has been observed even inside a genus. For example, *Phanerochaete chrysosporium* appeared to simultaneously degrade lignin along with all cell wall carbohydrates while *Phanerochaete carnosa* showed a selective (i.e. predominant) removal of lignin with mostly no loss of cellulose (Suzuki *et al.*, 2012). It is important to keep in mind that the ability of certain fungi to selectively degrade lignin is shaped by the environmental conditions, and therefore, under different conditions, the same fungi might produce different proportions of delignification. Still, the study of selective lignin degradation capabilities is interesting for biorefineries dedicated to the transformation of cellulose, as the removal of the lignin barrier improves the accessibility of the substrate for cellulases.

Besides wood decayers, additional lifestyles have been observed in Polyporales. Pathogens colonize living trees through wounds and cause wood rot disease (e.g. *Phaeolus schweinitzii* on conifers, *Wolfiporia coco* on pines, *Fomitopsis pinicola* on spruce, pines and birch, *Ganoderma boninense* on palm trees). Polyporales have also occasionally been detected as endophytes in the leaves and sapwood of trees, where they are suggested to stand as latent saprotrophs (Sokolski *et al.*, 2007; Kohler *et al.*, 2015). Finally, some have been described as potential temporary parasites of other fungi preliminarily established in the wood (Rayner *et al.*, 1987).

Polyporales use different sets of enzymes to degrade lignin. For example, within the phlebioid clade, some species have no laccase genes in their genome (*Phanerochaete chrysosporium*, *Phanerochaete chrysosporium* (Suzuki *et al.*, 2012), whereas four and five laccase genes have been identified in *Phlebia centrifuga* and *Phlebia radiata* respectively (Kuuskeri *et al.*, 2016). Similarly, in the genomes of core polyporoids, the numbers of lignin active POD vary significantly from one species to another. For example, no POD from the VP family were identified in the genome of *P. chrysosporium* whereas nine genes were found in that of *Polyporus brumalis,* of which five were secreted during growth on wood, responsible for its efficient delignification abilities (Miyauchi *et al.*, 2018).

Such a high diversity in enzymatic systems and enzyme gene repertoires in Polyporales is of high interest to biotechnologies aimed at the sustainable production of bio-based chemicals and

materials derived from recalcitrant plant tissues, such as wastes or co-products of agriculture and forestry, whose valorization does not compete with food or feed production. In particular, as indicated above, Polyporales fungi have developed efficient mechanisms for the oxidative degradation of non-starch polymers (Berrin *et al.*, 2017; Miyauchi *et al.*, 2018; Couturier *et al.*, 2018), which stimulates the screening of Polyporales strains for process development (Berrin *et al.*, 2012; Zhou *et al.*, 2015).
A recent phylogenomic study suggested that high diversification rates occurred in Polyporales (Varga *et al.*, 2019), which could have shaped diverse enzymatic mechanisms to adapt diverse ecological niches and substrates, and further strengthens the relevance of the exploration of this taxon for enzymatic diversity. Yet, further studies are required to extend our understanding on the genomic adaptations to wood decay and to assess the functional diversity of the enzymatic systems across Polyporales.

II.4. Regulation of the expression of fungal PCW degrading enzymes

Beyond PCWDE gene portfolios in fungal genomes, functional diversity may arise from the regulation of gene expression, at the transcriptional, translational or post-translational level.

In filamentous fungi, the production of PCWDE is largely regulated by the availability and abundance of carbon resources. In several ascomycetes fungi, cellulase and hemicellulase production is repressed when fungi are in presence of easily assimilable mono- or disaccharides, through carbon catabolism repression (CCR). On the contrary, the limitation of nutrient resources causes an increase in the expression of genes encoding ligninolytic enzymes (Alfaro *et al.*, 2020). In addition, other conditions such as the temperature, the pH, the presence of ions or reactive oxygen species contribute to the regulation of expression of PCWDE genes (Dashtban *et al.*, 2010). Several transcription factors have been identified which regulate the transcription of cellulase and hemicellulose genes in model ascomycete fungi. The deletion of the transcription factors ACE1 and ACE3 in *Trichoderma reesei* increased the expression of cellulase and xylanase genes respectively (Gupta *et al.*, 2016). Homologs for these transcription factors have been identified in the ascomycetes *Neurospora crassa* and in *Aspergilli* (Amore *et al.*, 2013). Recently, CCR has been shown to occur in the Basidiomycetes *Pleurotus ostreatus* and *Dichomitus squalens* (Daly *et al.*, 2019; Alfaro *et al.*, 2020). Yet, despite efforts for the identification of the transcription factors involved in CCR in Basidiomycota (Yoav *et al.*, 2018), they are still largely unknown.

However, several transcriptomic studies have shown that the transcription of PCWDE genes is finely regulated in Basidiomycetes during growth on biomass feedstocks. For example, in the

white-rot fungus *Pycnoporus coccineus*, a set of CAZymes were simultaneously expressed and secreted 3 days after transfer from a culture medium containing maltose to a medium containing lignocellulose as the sole carbon source. Among these enzymes, some target cellulose and β-1,4-glycans (AA9 LPMOs, GH3, GH5, GH6, GH7, GH45 and GH131). Others target hemicellulose; GH10s cleave the main chain of xylan, GH43s and GH51s are arabinofuranosidases and cleave the arabinose substitution of xylan, GH115 glucuronidases cleave the glucuronoyl substitution of woody xylan, CE4s and CE16s target acetyl xylans and CE15s target 4-O-Methyl glucuronoyl side chains of hemicellulose (Miyauchi *et al.*, 2017). The set of CAZyme induced genes was enriched in enzymes fused to a CBM1 carbohydrate binding module, that favours access of the catalytic module to insoluble substrates such as crystalline cellulose (Lehtiö *et al.*, 2003; Fong *et al.*, 2016). Noticeably, eight AA9 LPMO genes were strongly up-regulated, in a transcription profile that resembled that of the cellobiose dehydrogenase (CDH) gene, which suggested that the CDH might promote the activity of these LPMOs via electron transfer, as suggested from in vitro enzymatic assays (Bissaro *et al.*, 2017). The induction of transcription of this set of genes further increased after 7 day-growth on lignocellulosic substrates. On the contrary, the genes coding for lignin-active PODs were strongly induced at day 3, and their transcript level decreased at day 7. Surprisingly, the enzymes were only detected at day 7 in the culture medium. A similar delay in the secretion of POD enzymes was observed in the white-rot fungi *Phanerochaete carnosa* and *Ceriporiopsis subvermispora* (MacDonald & Master, 2012; Hori *et al.*, 2014). It was suggested that the tight regulation of the expression of these peroxidases might deconstruct the lignin polymers and thereby facilitate access of cellulolytic and hemicellulolytic enzymes to their substrate while preventing damaging to the hyphae by oxidized compounds. The tuned regulation of PCWDE gene transcription and protein expression was also elegantly shown on miniaturized wood colonization assays with the brown-rot fungus *Postia placenta* (Zhang *et al.*, 2016). The authors observed the generation of ROS at the hyphal front whereas hydrolytic enzymes were produced at later stages during the hyphal progression in the wood. Uncoupling oxidative degradation of lignin from hydrolytic degradation of polysaccharides likely resulted in an oxidative pretreatment of lignocellulose and the protection of the hydrolytic enzymes from oxidative damages. Interestingly, a similar sequential induction of oxidative and hydrolytic mechanisms was observed by Navarro *et al.*, (2014), who showed that AA9 LPMOs, which cleave cellulose by oxidation, were produced before hydrolytic cellobiohydrolases when the soil inhabiting agaricomycete fungus *Laetisaria arvalis* was grown on crystalline cellulose.

Several studies also highlighted different gene regulations when fungi are grown on different woody substrates. For example, Daly *et al.*, (2018) showed that the white-rot fungus *Dichomitus squalens* produced more mannolytic enzymes during growth on softwood (higher mannan content) and more xylanolytic enzymes during growth on hardwood (higher xylan content). In addition, lignin-active MnPs were more abundantly produced during growth on softwood

(which contains guaiacyl (G) lignin) than on hardwood (which contains syringyl/guaiacyl (S/G) lignin). These findings were in line with the acquisition of MnPs by white-rot fungi for adaptation to gymnosperm woods (Floudas *et al.*, 2012) and with MnPs having a higher efficiency for the modification of G-lignin compared to S-lignin (MacDonald *et al.*, 2016).

II.5. Fungi as a source of biocatalysts for synthetic biology: fungal terpene synthases

II.5.1. The roles of terpenes in fungi

The origin of the name "terpene" is "turpentine", which is the resin of trees; the viscous substance that the plants secrete in response to injury for their protection against wounds, pathogens and insects. The volatile fraction of resin is composed of a mixture of terpenes. Although terpenes are mainly present and extensively studied in plants, they are ubiquitous and involved in a wide range of biological functions, in which they have basic roles as membrane components (steroids), photosynthetic pigments (carotenoids), or terpenoid-derived hormones (in plants and animals), but also specialized functions in chemical interactions or in protection against abiotic and biotic environmental stresses. Fungi produce a wealth of such compounds that are not essential for their survival and growth, but that act as crucial intermediates at the front line of their ecology (Rokas *et al.*, 2020). They ensure the interaction that fungi have with other organisms sharing the same ecological niches or competing for the carbon source. These interactions include defense, protection, attraction of beneficial organisms and repulsion of harmful ones (El Ariebi *et al.*, 2016; O'Leary *et al.*, 2019). Except for a few ascomycete model species (Keller *et al.*, 2005), there is little information currently available regarding the terpenes produced by fungi, and it is expected that a huge diversity of terpenes can be discovered from fungi having different lifestyles or living in different ecological niches.

II.5.2. Terpene industrial uses and natural functions

Terpenes belong to the largest and most chemically and structurally diverse class of natural products, having more than 80000 different structures currently described. The physico-chemical properties and biological activities of these natural products have attracted many flavor, perfum and bioenergy industries. Also, several of these compounds have led to the successful development of pharmaceuticals, such as the well-known anticancer agent TAXOL®(paclitaxel) and the drug artemisinin that has been used for more than 2000 years to treat malaria (Table3) (Tabima *et al.*, 2020). The industrial access to known terpenes still relies

on their extraction from natural resources such as cultivated plants (mint, camphor, etc.), an approach that is in contradiction with biodiversity protection and sustainable development. Another problem in the access of those valuable molecules comes from their structural complexity which makes their chemical synthesis very tedious and prohibitive from an economic point of view. As consequence, the industrial access to terpenes must move from current natural extractions and chemical synthesis to eco-friendly synthetic biology.

Table 3: Terpenoids from higher fungi and their bioactivities. Table from Xiao & Zhong, 2016.

Terpenoids	Higher Fungi	Bioactivities
Sesquiterpenoids		
Illudins	*Omphalotus olearius*	Antitumor, antimicrobial
3α,6β-Dihydroxycinnamolide	*Inonotus rickii*	Anticancer
Hirsutane-type	*Stereum hirsutum*	Antimicrobial and antitumor
Enokipodins C–D, E–J	*Flammulina velutipes*	Antimicrobial, Antifungal
Nambinones A–C	*Neonothopanus nambi*	Anticancer
Diterpenoids		
Pleuromutilin	*Clitopilus passeckerianus*	Antimicrobial
Secoscabronine M	*Sarcodon scabrosus*	Anticancer
Striatoids A–F	Cyathus striatus	Neurotrophic activity
Cyathins D–H	*Cyathus africanus*	Anti-inflammatory and cytotoxic
Neosarcodonin A–C	*Sarcodon scabrosus*	Anti-inflammatory
Triterpenoids		
Ganoderic acids	*Ganoderma lucidum*	Antitumor, anti-HIV, antimicrobial, antimetastasis, antioxidation, etc.
Lucidenic acids	*G. lucidum*	Anticancer
Lanostane-type	*Naematoloma fasciculare*	Anticancer
Ganoboninketals A–C	*Ganoderma boninense*	Anticancer
Cattienoids A–C	*Tomophagus cattienensis*	Antitumor

II.5.3. Terpene biosynthesis

In fungi the biosynthesis of terpenes starts with the mevalonate pathway that uses 18 enzymes to transform glucose into the two precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) (Figure 10). The head-to-tail fusion of the C-5 DMAPP with one, two, or three units of C-5 IPP form respectively, the linear geranyl diphosphate (GPP, C10, monoterpenoids), the linear farnesyl diphosphate (FPP, C15, sesquiterpenoids) and the linear geranylgeranyl diphosphate (GGPP, C20, diterpenoids) (Christianson D.W., 2008). Next, these linear precursors of terpenes are cyclized by terpene synthases into polycyclic structures with numerous stereogenic centers (Figure 11).

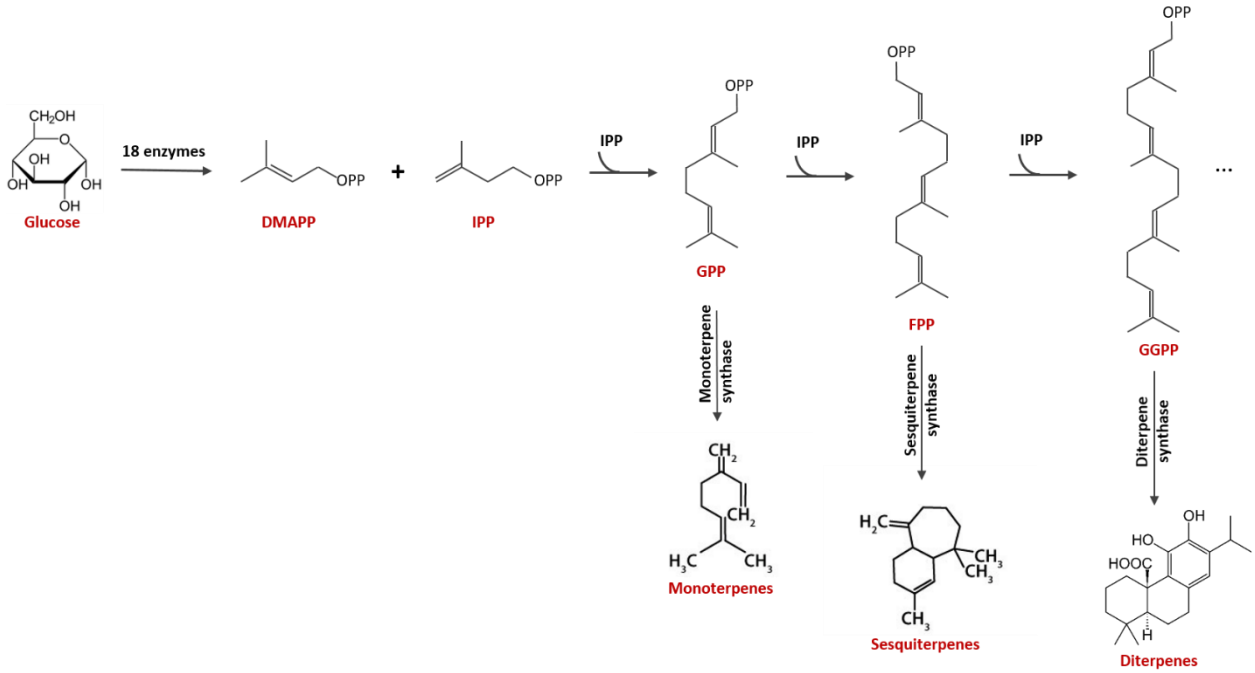


Figure 10: Biosynthesis of terpenes in fungi.

The diverse terpenoid compounds obtained, are further modified by hydroxylation, glycosylation, methylation, acylation or peroxidation. These chemical modifications are done by tailoring enzymes such as Cytochrome P450s (CYPs), alcohol dehydrogenases, acyl transferases and glycosyl transferases, that are typically co-located in the genomes with a terpene cyclase gene as part of a biosynthesis gene cluster. In Ascomycota and Basidiomycota, sesquiterpene cyclases are more abundant than mono- and di-terpene cyclases (Schmidt-Dannert C., 2015).

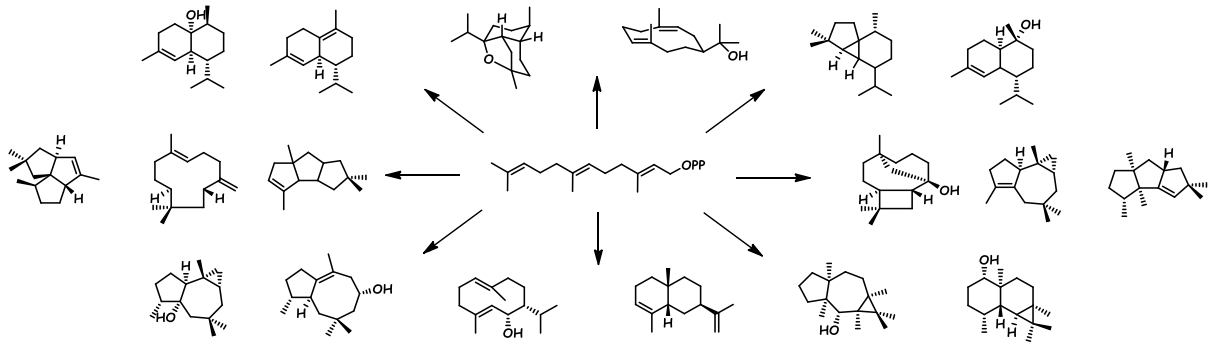


Figure 11: Some of the various naturally occurring polycyclic sesquiterpene structures resulting from the cyclization of farnesyl diphosphate.

II.5.4. Sesquiterpene cyclases

Sesquiterpene synthases (STS) are class I terpene synthases responsible for cycling the FPP ((2E,6E)-FPP), to form thousands of structurally diverse sesquiterpenoids. Some STSs catalyse the formation of a single product (Proctor & Hohn, 1993), while others form multiple products (Agger *et al.*, 2009; Drew *et al.*, 2016; Wu *et al.*, 2016). The cyclisation happens by cleaving the pyrophosphate group of FPP at the entrance of the active site via a metal ion co-factor, usually Mg^{2+}, which is coordinated by two conserved aspartate-rich motifs, DDXXD/E and NSE/DTE (Christianson D.W., 2008). The elimination of the pyrophosphate group generates a primary carbocation from (E,E)-FPP that can undergo two different ring closures (1,10 or 1,11) or is isomerized to a secondary carbocation from (3R)-NPP, which can undergo four different ring closures (1,6; 1,7; 1,10 or 1,11) (Figure 12). Subsequently, the carbocation undergoes a series of further cyclizations, ring rearrangements, methyl, and hydride shifts, until a final deprotonation. This deprotonation is mediated either by a water molecule attack (Pinedo *et al.*, 2008) or through proton abstraction (Chen *et al.*, 2013) resulting in the release of the sesquiterpene from the enzyme active site.

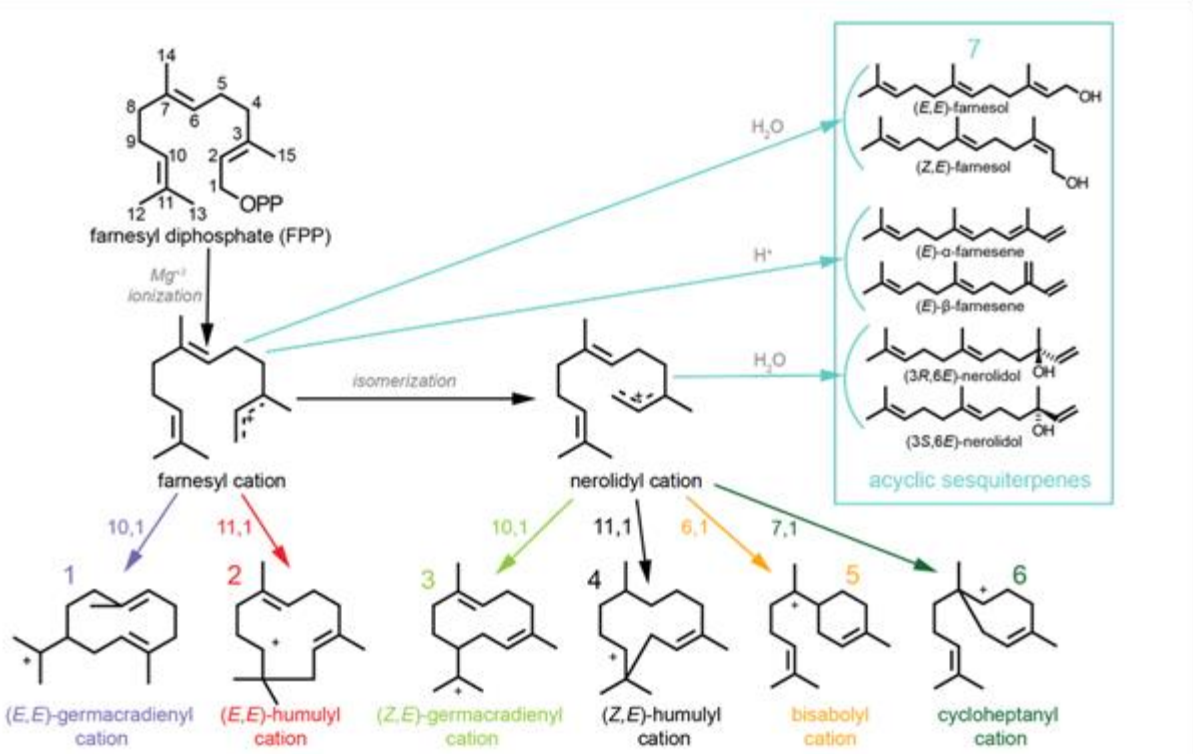


Figure 12: Sesquiterpene cyclization scheme. Different sesquiterpene synthases can catalyse distinct carbocation cyclization mechanisms: the enzyme can catalyse a 1,10 or 1,11 cyclization following removal of PPi from (2E,6E)-FPP, yielding a E,E-germacradienyl cation or trans-humulyl cation, respectively. Alternatively, the enzyme can cause (2E,6E)-FPP to undergo an initial isomerization yielding (3R)-nerolidol pyrophosphate (NPP), followed by a 1,6 or 1,10 cyclization yielding a bisabolyl cation or Z,E-germacradienyl cation, respectively. Modified from (Durairaj *et al.*, 2019).

Despite those two well conserved motifs, STS show relatively low level of conservation in their protein sequences.

To the best of our knowledge, about 100 fungal STS have been characterized to date. In ascomycetes, efforts have been focused on enzymes involved in the synthesis of mycotoxins, such as DON (deoxynivalenol) (e.g. trichodiene synthase from *Fusarium graminearum* (Proctor *et al.*, 1995)), trichodermin (e.g. trichodiene synthase from Trichoderma brevicompactum (Tijerino *et al.*, 2011)), the PR toxin (aristolochene synthase from *Penicillium roqueforti* (Cheeseman *et al.*, 2014)), or on fungi of agriculture importance (e.g. *Colletotrichum acutatum* and *Botrytis cinerea* (Wang *et al.*, 2009; Buchvaldt Amby *et al.*, 2016)). In addition, a few STS have been characterized from endophyte fungi, which attract interest for secondary metabolites that have toxicity against insect pests (Wu *et al.*, 2016). In basidiomycetes, recent efforts have been largely dedicated to the search of STS involved in the synthesis of Δ6-protoilludene, a precursor of anticancer molecules, for example in edible fungi (e.g. *Agrocybe aegerita* (Zhang *et al.*, 2020), in tree pathogens (*Armillaria gallica* (Engels *et al.*, 2011) and

Heterobasidion annosum (Zhang *et al.*, 2020)) and in wood decay fungi (e.g. *Omphalotus olearius* (Wawrzyn *et al.*, 2012), *Postia placenta* (Ichinose & Kitaoka, 2018) or *Stereum hirsutum* (Nagamine *et al.*, 2019), among others.

Interestingly, phylogenetic analyses of the characterized STS from basidiomycetes have suggested that STS with similar cyclization mechanism cluster together. From these studies, a classification of fungal STS in four clades has been proposed; clade I STS catalyze the 1,10-cyclization of (2E,6E)-FPP carbocation, clade II STS catalyze the 1,10-cyclization of the (3R)-NPP carbocation; clade III STS catalyze the 1,11 cyclization of (2E,6E)-FPP carbocation (trans-humulyl-type cyclases) and clade IV STS catalyze the 1,6 or 1,7 cyclization of the (3R)-NPP carbocation (Figure 13), (Wawrzyn *et al.*, 2012).

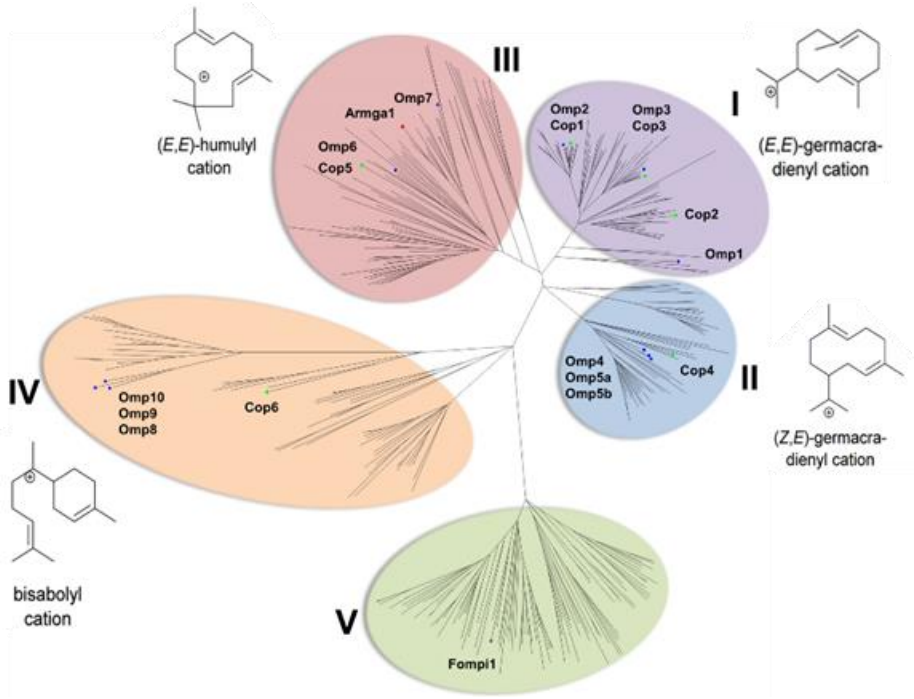


Figure 13: Unrooted Neighbor-Joining phylogram of sesquiterpene cyclase homologs identified in 42 Basidiomycota genomes. Five clades are highlighted with colors. The characterization of a few enzymes suggested that the trans-humulyl-type cyclases (clade III), the sesquiterpene cyclases responsible for 1,6 or 1,7 cyclization of (3R)-NPP (clade IV), for 1,10 cyclization of either (E,E)-FPP (clade I) or (3R)-NPP (clade II) group together. Modified from Wawrzyn *et al.*, 2012.

II.5.5. Current bioinformatic tools for the gene annotation of terpene cyclases

The identification of fungal STS can be done either by a blast search of available characterized enzymes, by searching for conserved Pfam domains related to terpenes (PF01397, PF03936), by using available tools for the identification of terpene cyclases, such as Terzyme (Priya *et al.*, 2018), or for the identification of gene clusters for the biosynthesis secondary metabolites, such as SMURF (Khaldi *et al.*, 2010), or antiSMASH (Medema *et al.*, 2011). However, these tools are mostly specialized in the analysis of Non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) (Weber & Kim, 2016), or based on Pfam domains that were built from plant or bacterial sequences and might not match the sequence polymorphism found in fungal terpene cyclases,. Thus, optimized bioinformatic tools should be developed in order to discover new fungal terpene cyclases from the wealth of fungal genomes now available.

Chapter III THESIS OBJECTIVES

Wood-decay fungi are among the most important microorganisms in forest ecosystems where they play a crucial role in carbon release. They have evolved for over 290 million years to become most efficient in utilizing the available resources in their ecological niches. Consequently, they have acquired diverse enzymatic mechanisms to degrade the complex biopolymers and access carbon and other nutrients. On a biotechnological point of view, utilizing fungal enzymatic toolkits or getting inspired by the enzymatic systems fungi have used to adapt to their natural habitats could help improving the design of biocatalysts aimed at the use of biopolymers as a renewable carbon source. Despite many discoveries on the enzymatic machineries involved in wood decomposition, the vision on the evolutionary adaptation to wood decay and genome diversity remains incomplete. Beyond lignocellulose modification and degradation, the extraordinary biodiversity within the kingdom Fungi is the cornerstone of unraveling the enzymatic functions at play during the colonization of their ecological niches.

The first and second section of my thesis results aim at exploring the functional and genomic diversity among wood decayers at different taxonomical levels. The first section covers a single genus, *Pycnoporus,* comprising four white-rot species encountered in different geoclimatic areas. The main goals of this study were to assess the intra-genus diversity and to identify the conserved enzymatic responses to diverse lignocellulose substrates, by comparing the genomes, transcriptomes and secretomes of the four species. The second section covers a taxonomic order, the order Polyporales, which comprises highly efficient wood decay fungi. The objective of this study was to identify the evolutionary histories that shaped the repertoires of genes involved in wood decay in this order.

The third section of the results goes beyond the lignocellulose degrading enzymatic machineries. Because fungi share the same ecological niches with other microorganisms, they are facing a constant competition for the resources. Therefore, they have acquired strategies to communicate and interact with other microorganisms or with their host plants, and to defend and protect themselves in harsh environmental conditions. Those strategies include the ability to produce a diversity of secondary metabolites such as terpenes. Despite the tremendous terpene diversity fungi are believed to produce, the discovery of novel terpenes from the fungal kingdom remains very slow, due to the lack of available tools for mining fungal genomes for terpene synthases. Therefore, the third section aimed at developing a new tool for mining terpene synthase genes in fungal genomes, with an attempt to predict the cyclisation mechanisms that might lead to the final terpene products.

During my PhD, I also had the opportunity to contribute to a phylum-wide analysis of the methionine sulfoxide reduction system in Fungi, to the analysis of the genomic features related to the selective degradation of lignin over polysaccharides by the fungus *Polyporus brumalis* and to the study of a coupled enzymatic system of peroxidase-copper radical oxidase in the plant pathogen *Colletotrichum* sp.. These works are presented in annex I, II and III.

Chapter IV RESULTS

IV.1. Conserved white-rot enzymatic mechanisms for wood decay in the Basidiomycota genus Pycnoporus

Chapter IV.1 summary

The dynamics of deadwood decay in forests contribute to the carbon balance and is related to successions of fungal communities; notably wood decayers such as white-rot fungi that have the power to degrade all plant cell wall polymers, including lignin and polysaccharides. The pivotal role white-rot fungi play in forest ecosystems has stimulated research efforts to understand the enzymatic mechanisms involved in wood degradation. Independent studies on individual species have highlighted common trends in the production of lignin-, cellulose-, hemicellulose- and pectin-degrading enzymes during white-rot fungi growth on diverse plant biomasses. So far, those different studies have been largely based on diverse fungal taxa, thus providing an overview of the enzymatic mechanism employed during lignocellulose breakdown. However, intra-genus diversity of the enzymatic toolbox has not been extensively explored in white-rot fungi.

In this study, we aimed at exploring the functional and genomic diversity related to lignocellulose breakdown at the genus level, using the genus *Pycnoporus,* which is renowned for its great potential for novel biocatalysts and comprises four white-rot species encountered in different geographical area..

To investigate the *Pycnoporus* intra-genus genomic and functional diversity on one hand and to identify the core enzymatic responses to diverse lignocellulose substrates on the other hand, the genomes of the four *Pycnoporus* species were sequenced and used for comparative genomics, transcriptomics and secretomics. In addition, a methodology for cross-species comparison of the gene expression profiles was developed to identify a conserved set of genes responsive to lignocellulosic substrates that outlines key enzymatic mechanisms for wood decomposition among those fungi.

Despite the species coming from geographically distant areas, the sequenced genomes did not show any major rearrangements in structure and strong conservation in protein coding gene composition. In particular, we observed strong conservation of the CAZomes, confirming poor genomic diversity at the genus level. Conversely, the species showed surprisingly different abilities to grow on (ligno)cellulosic substrates *in vitro*, indicating that the gene repertoires obtained from genome sequencing are not sufficient to predict the ability of a strain to grow on recalcitrant raw biomass.

This work is presented in the manuscript “Conserved white rot enzymatic mechanism for wood decay in the Basidiomycota genus Pycnoporus”, by Shingo Miyauchi*, Hayat Hage*, Elodie Drula, Laurence Lesage-Meessen, Jean-Guy Berrin, David Navarro, Anne Favel, Delphine Chaduli,

Sacha Grisel, Mireille Haon, François Piumi, Anthony Levasseur, Anne Lomascolo, Steven Ahrendt, Kerrie Barry, Kurt M. LaButti, Didier Chevret, Chris Daum, Jérôme Mariette, Christophe Klopp, Daniel Cullen, Ronald P. de Vries, Allen C. Gathman, Matthieu Hainaut, Bernard Henrissat, Kristiina S. Hildén, Ursula Kües, Walt Lilly, Anna Lipzen, Miia R. Mäkelä, Angel T. Martinez, Mélanie Morel-Rouhier, Emmanuelle Morin, Jasmyn Pangilinan, Arthur F.J. Ram, Han A.B. Wösten, Francisco J. Ruiz-Dueñas, Robert Riley, Eric Record, Igor V. Grigoriev and Marie-Noëlle Rosso,' published in DNA Research, 2020;27(2):dsaa011. doi: 10.1093/dnares/dsaa011.

DNA Research, 2020, 00(0), 1–14
doi: 10.1093/dnares/dsaa011
Advance Access Publication Date: 12 June 2020
Full Paper

OXFORD

Full Paper

Conserved white-rot enzymatic mechanism for wood decay in the Basidiomycota genus *Pycnoporus*

Shingo Miyauchi[1,2,†,§], Hayat Hage[1,§], Elodie Drula[1], Laurence Lesage-Meessen[1,3], Jean-Guy Berrin[1], David Navarro[1,3], Anne Favel[1,3], Delphine Chaduli[1,3], Sacha Grisel[1], Mireille Haon[1], François Piumi[1,‡], Anthony Levasseur[1,¶], Anne Lomascolo[1], Steven Ahrendt[4], Kerrie Barry[4], Kurt M. LaButti[4], Didier Chevret[5], Chris Daum[4], Jérôme Mariette[6], Christophe Klopp[6], Daniel Cullen[7], Ronald P. de Vries[8,9], Allen C. Gathman[10], Matthieu Hainaut[11,12], Bernard Henrissat[11,12], Kristiina S. Hildén[9], Ursula Kües[13,14], Walt Lilly[10], Anna Lipzen[4], Miia R. Mäkelä[9], Angel T. Martinez [15], Mélanie Morel-Rouhier[2], Emmanuelle Morin[2], Jasmyn Pangilinan[4], Arthur F.J. Ram[16], Han A.B. Wösten[17], Francisco J. Ruiz-Dueñas[15], Robert Riley[4], Eric Record[1], Igor V. Grigoriev[4,18], and Marie-Noëlle Rosso [1,*]

[1]INRAE, UMR1163, Biodiversity and Biotechnology of Fungi, Aix Marseille University, 13009 Marseille, France, [2]INRAE, UMR1136, Interactions Arbres/Microorganismes, Université de Lorraine, Nancy, France, [3]INRAE, CIRM-CF, UMR1163, Aix Marseille University, Marseille, France, [4]US Department of Energy, Joint Genome Institute, Walnut Creek, CA, USA, [5]INRAE, UMR1319, Micalis, Plateforme d'Analyse Protéomique de Paris Sud-Ouest, Jouy-en-Josas, France, [6]INRAE, Genotoul Bioinfo, UR875, Mathématiques et Informatique Appliquées de Toulouse, Castanet-Tolosan, France, [7]USDA Forest Products Laboratory, Madison, WI, USA, [8]Fungal Physiology, Westerdijk Fungal Biodiversity Institute and Fungal Molecular Physiology, Utrecht University, Utrecht, The Netherlands, [9]Department of Microbiology, University of Helsinki, Helsinki, Finland, [10]Department of Biology, Southeast Missouri State University, Cape Girardeau, MI, USA, [11]CNRS, UMR7257, AFMB, Aix Marseille University, Marseille, France, [12]INRAE, USC1408, AFMB, Marseille, France, [13]Department of Molecular Wood Biotechnology and Technical Mycology, Büsgen-Institute, Georg-August-University Göttingen, Göttingen, Germany, [14]Center for Molecular Biosciences (GZMB), Georg-August-University Göttingen, Göttingen, Germany, [15]Centro de Investigaciones Biológicas, CSIC, Madrid, Spain, [16]Molecular Microbiology and Biotechnology, Institute of Biology Leiden, Leiden University, Leiden, The Netherlands, [17]Microbiology, Utrecht University, Utrecht, The Netherlands, and [18]Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA, USA

[†]Present address: Department of Plant Microbe Interactions, Max Planck Institute for Plant Breeding Research, Köln, Germany.

[‡]Present address: UMR INRAE, BDR, ENVA, Université Paris Saclay, 78350 Jouy en Josas, France.

[¶]Present address: IRD, Microbes, Evolution, Phylogeny and Infection, APHM, IHU Méditerranée Infection, Aix Marseille University, Marseille, France.

[§]These authors are considered co-first authors.

*To whom correspondence should be addressed. Tel. +33 491828607. Fax +33 491828601.
Email: marie-noelle.rosso@inrae.fr

Received 18 March 2020; Accepted 5 June 2020

Abstract

White-rot (WR) fungi are pivotal decomposers of dead organic matter in forest ecosystems and typically use a large array of hydrolytic and oxidative enzymes to deconstruct lignocellulose. However, the extent of lignin and cellulose degradation may vary between species and wood type. Here, we combined comparative genomics, transcriptomics and secretome proteomics to identify conserved enzymatic signatures at the onset of wood-decaying activity within the Basidiomycota genus *Pycnoporus*. We observed a strong conservation in the genome structures and the repertoires of protein-coding genes across the four *Pycnoporus* species described to date, despite the species having distinct geographic distributions. We further analysed the early response of *P. cinnabarinus*, *P. coccineus* and *P. sanguineus* to diverse (ligno)-cellulosic substrates. We identified a conserved set of enzymes mobilized by the three species for breaking down cellulose, hemicellulose and pectin. The co-occurrence in the exo-proteomes of H_2O_2-producing enzymes with H_2O_2-consuming enzymes was a common feature of the three species, although each enzymatic partner displayed independent transcriptional regulation. Finally, cellobiose dehydrogenase-coding genes were systematically co-regulated with at least one AA9 lytic polysaccharide monooxygenase gene, indicative of enzymatic synergy *in vivo*. This study highlights a conserved core white-rot fungal enzymatic mechanism behind the wood-decaying process.

Key words: wood decay, lignocellulose, CAZyme, lytic polysaccharide monooxygenase, Class II Peroxidase

1. Introduction

Saprotrophic fungi of Northern Hemisphere and tropical forests impact the carbon cycling through mineralization and alteration of C storage in wood and litter dead organic matter.[1,2] White-rot (WR) fungi are wood decayers with the capacity to mineralize lignin with ultimate formation of CO_2 and H_2O.[3]. WR fungi deploy a wide arsenal of hydrolytic and oxidative enzymes to degrade wood and their genomes typically contain genes coding for glycoside hydrolases, carbohydrate esterases and polysaccharide lyases that collectively cleave cellulose, hemicellulose and pectin backbones and lateral chains, and oxidative enzymes that target the highly recalcitrant lignin, crystalline cellulose or cellulose-bound xylan.[4,5] Beyond these shared features, several studies have highlighted significant polymorphism between WR fungi regarding their ability to selectively degrade lignin over cellulose[6–8] and in the gene portfolios involved in lignocellulose breakdown.[9–11] Scarce studies at the intra-genus level have shown that functional diversity between species may arise from diversity in gene content.[12,13]

Among WR fungi, the genus *Pycnoporus* (Basidiomycota, Agaricomycetes) has been studied for the efficiency of lignin degradation, the capacity to secrete laccases and biotechnological applications related to aromatic compound functionalization, biopolymer synthesis and biomass pre-treatment in the pulp and paper industry.[14] Four *Pycnoporus* species have been differentiated,[15,16] which form a monophyletic group within the Trametes clade.[17] The four species are found in different geo-climatic areas with limited geographical overlap; *P. cinnabarinus* is widely distributed in the Northern hemisphere, *P. coccineus* is found in countries bordering the Indian and Pacific Oceans, *P. sanguineus* is found in the tropics and subtropics of both hemispheres and *P. puniceus* is found in paleotropical areas.[15,16] The four species are found on stumps and either standing or fallen trunks of deciduous trees.

Our aim was to investigate the *Pycnoporus* intra-genus genomic and functional diversity focusing on lignocellulose breakdown. We examined whether divergence in distinct geographic areas had led to genomic diversity, and if there was a signature of conserved enzymatic mechanics in terms of transcriptome and exo-proteome responses to lignocellulosic substrates. We sequenced the genomes of *P. coccineus*, *P. sanguineus* and *P. puniceus* monokaryotic strains. We overviewed the genomic features among the three species, in comparison to the previously sequenced genome of *P. cinnabarinus*[9] and to other evolutionarily related wood-decay fungi. Then, we captured the transcriptomic and exoproteomic responses of three *Pycnoporus* species to a panel of cellulosic and lignocellulosic substrates representative of Gramineae and hardwoods. The focus was the early responses to the substrates in order to minimize inter-species differences influenced by varied growth abilities on the substrates. Our omics integrative approach enabled to identify a common set of lignocellulose-degrading enzymes mobilized by the fungi at the initial stage of lignocellulose degradation, leading to discoveries of genus-wide conserved expression patterns indicative of conserved enzymatic synergies.

2. Materials and methods

2.1. Genome sequencing and assembly

The monokaryotic strains *P. coccineus* BRFM 310, *P. sanguineus* BRFM 1264 and *P. puniceus* BRFM 1868 were generated after fruiting of the parental strains BRFM 66 (IMB WOO6-2), BRFM 902 and BRFM 1856, respectively, as described previously[14] (Supplementary Information). All strains were maintained at the International Centre of Microbial Resources (CIRM; https://www6.inra.fr/cirm/). The *P. coccineus* BRFM 310 genome was sequenced using the Illumina platform (99.4×) and assembled with AllPathsLG

version R46652[18] (GenBank accession number: NCSW00000000). The *P. sanguineus* BRFM 1264 genome was sequenced using 454 (16.8×) and Solexa (87×) technologies and assembled with CABOG[19] (GenBank accession number: VOCM00000000). The *P. puniceus* BRFM 1868 genome was sequenced using PacBio technology (97×) and assembled with FALCON, improved with finisherSC,[20] polished with Quiver (GenBank accession number: VICQ00000000). The three genomes were annotated using the JGI annotation pipeline,[21] which takes multiple inputs (scaffolds, ESTs and known genes) and runs several analytical tools for gene prediction and annotation, and deposits the results in the JGI Genome Portal MycoCosm (http://genome.jgi.doe.gov/fungi). The previously sequenced and annotated genome of *P. cinnabarinus* BRFM 137[9] was also deposited in MycoCosm.

2.2. Construction of phylogenetic tree

We constructed a phylogeny based on orthologous genes among the selected fungi using FastOrtho with the parameters set to 50% identity, 50% coverage and inflation 3.0.[22] The protein sequences used for the process were downloaded from the JGI fungal portal MycoCosm. We identified the clusters of orthologous genes with single copy genes, aligned the sequences of each cluster with MAFFT 7.221,[23] eliminated ambiguous regions (containing gaps and poorly aligned regions) and concatenated the alignments with Gblocks 0.91b.[24] We constructed a phylogenetic tree with RAxML 7.7.2[25] using the standard algorithm, the PROTGAMMAWAG model of sequence evolution and 500 bootstrap replicates.

2.3. Comparative genomic analysis

Genome completeness with single copy orthologues was calculated using BUSCO v3.0.2 with default parameters.[26] The coverage of transposable elements (TEs) in genomes was calculated using a custom pipeline transposon identification nominative genome overview.[27] The counts for plant cell wall (PCW)-degrading enzymes, predicted secreted auxiliary activity (AA) enzymes and fungal cell wall (FCW)-degrading enzymes were combined and visualized with custom R scripts, proteomic information navigated genomic outlook[28] incorporating R packages ggplot2, ggtree and egg.[29–31]

2.4. Expert functional annotations

Genes from the *A*- and *B*-mating type loci were identified and manually curated as described in Kues et al.[32] CAZymes and AA were annotated as in Lombard et al.[33] Gene models from AA2 Class II peroxidases, AA3_2 glucose-methanol-choline (GMC)-oxidoreductases and AA5 copper radical oxidases were further inspected by sequence-by-sequence exhaustive analysis and phylogenetic analysis. Peptidases were annotated using Blastx searches of gene models against InterPro and MEROPS databases[34,35] followed by manual curation. Glutathione transferases (GST)-coding genes were annotated with a combination of automated blastp using functionally characterized GST sequences from *P. chrysosporium*,[36–38] phylogenetic analysis and active site comparison. Genes coding for hydrophobins, laccases, the secretory pathway and carbon catabolism were also manually inspected. Proteins were predicted secreted if they fulfilled three conditions: (i) presence of a secretion signal peptide, (ii) absence of endoplasmic reticulum retention motif and (iii) absence of transmembrane helix outside the signal peptide, as previously described.[39] Predicted secreted proteins shorter than 300 amino acids were annotated as small secreted proteins (SSPs). Detailed analyses of expert annotations are provided in the Supporting Information file.

2.5. Cultures

Media for cultures on agar plates contained diammonium tartrate (1.84 g/l), yeast nitrogen base (0.17 g/l), agar (15 g/l) and were supplemented with either maltose (20 g/l), Avicel (AVI) PH 101 (Fluka) (15 g/l), ground and sifted wheat straw (WS) fragments <2 mm (15 g/l), *Pinus halepensis* pine wood fragments <2 mm (15 g/l) or *Populus tremuloides* Wiley-milled aspen (Asp) fragments (180 µm < fragment size < 2 mm; 15 g/l). The plates were inoculated with one fungal disk (4 mm diameter) of 7-day-old mycelia and incubated at 30 °C. Liquid cultures were maintained at 30 °C in a rotary shaker at 120 rpm in 250-ml Erlenmeyer flasks containing 100 ml of culture medium (Supplementary Information) supplemented with either maltose (20 g/l), AVI (15 g/l), WS fragments (15 g/l), *P. halepensis* wood fragments (15 g/l) or *P. tremuloides* fragments (15 g/l). Each culture was done in triplicate. Inoculums of the liquid cultivations were prepared as described in Herpoël et al.[40]

2.6. Integration of transcriptome and exo-proteome profiles

LC-MS/MS analysis of the secreted proteins was performed as described in Navarro et al.[41] Briefly, 10 µg of diafiltered proteins were loaded on SDS-PAGE gels and allowed to migrate on a 0.5 cm length. Each lane was cut into two slices for in-gel digestion according to a standard trypsinolysis protocol. On-line analysis of the peptides was performed with a Q-exactive mass spectrometer (Thermo Fisher Scientific), using a nanoelectrospray ion source. Protein identification was performed by querying MS/MS data against the genome of *P. cinnabarinus* BRFM 137, *P. coccineus* BRFM 310 or *P. sanguineus* BRFM 1264, together with an in-house contaminant database, using the *X*! Tandem Cyclone software. All peptides that matched with an *E* value lower than 0.05 were parsed with *X*! Tandem pipeline software. Proteins identified with at least two unique peptides and a log (*E* value) lower than −2.6 were validated.

For total RNA extractions, mycelia were ground in liquid nitrogen using a Freezer/Mill Cryogenic Grinder (SPEX Sample Prep, UK). Total RNA was extracted from 100 mg ground tissue in 1 ml TRIZOL (Ambion). Nucleic acids were precipitated with isopropanol, resuspended in water and treated with RNase-Free DNase I (QIAGEN). Total RNA was precipitated with LiCl and resuspended in DEPC-treated water. RNA quantity and quality were determined using the Experion RNA StdSens kit (QIAGEN). The transcriptome response of *P. sanguineus* to pine could not be analysed because of poor quality of the extracted RNAs. Double stranded cDNAs were synthesized from PolyA RNA and fragmented (200–300 bp) before construction of the sequencing libraries (Kapa Library Amplification Kit; Kapa Biosystems). Sequencing was done on the Illumina HighSeq-2500 JGI platform generating paired end reads of 150 bp each. Paired end 150 bp Illumina reads were trimmed for quality and aligned to the corresponding genome using TopHat 2 with only unique mapping allowed.[42] Gene models for which the mean raw read counts were inferior to 5 were considered as not transcribed and their read counts were changed to 0.

The counts of mapped Illumina reads from biological triplicates of each growth condition (GEO accession number GSE82486) were normalized with the DESeq2 package and log2 transformed.[43] The normalized read counts of genes coding for CAZymes, peptidases, hydrophobins and SSPs from *P. cinnabarinus* BRFM 137, *P.*

coccineus BRFM 310 and *P. sanguineus* BRFM 1264 were retrieved and combined by conducting; (i) removal of batch effects with Combat function in SVA package[44] and (ii) quantile normalization with the preprocessCore package.[45] We used self-organizing map (SOM) to group genes into nodes according to similar transcript levels obtained from the different substrate conditions. Self-organising maps were constructed with the R package kohonen.[46] The genes showing similar transcription levels were sorted and grouped into nodes of SOMs. It was empirically found that about 35 genes in a single node of the SOM gave the best resolution of the gene clusters. In terms of the standard formula '$X \times \text{sqrt}(N)$' to calculate the number of map units, where N was the number of the rows/genes of the data, X was 1.5. The number of iterations (epochs) was 100 times more than the map units to minimize the mean distance between the weights of the neighbouring nodes. The default initialization, learning rate and radius were used. Hexagonal SOM models were constructed. The mean reads (>12 log2) of the nodes (grouped genes) with the replicates combined were calculated for each substrate.

We integrated SOM with the protein secretome information analysed by peptide-LC-MS/MS, using SOM Harboring Informative Nodes with Gene Ontology.[10,47,48]

2.7. Transcription regulation of co-orthologous genes

One-to-one orthologous genes from *P. cinnabarinus* BRFM 137, *P. coccineus* BRFM 310 and *P. sanguineus* BRFM 1264 were retrieved using OrthoFinder v. 2.3.8.[49] Heatmaps were created on the log2-fold change of transcript read counts in each growth condition when compared with growth on maltose after DESeq2 normalization using the 'Heatmap' function from the package 'ComplexHeatmap' v1.10.1 in R. Pairwise comparisons of gene expression based on Pearson correlation coefficients among all replicates was performed on CAZyme, peptidase, hydrophobin and SSP co-orthologs after read count normalization by DESeq2, batch effect removal and quantile normalization 'cor' function in R and visualized as heatmap with R package, ggplot2.

3. Results and discussion

3.1. Four *Pycnoporus* species share similar genomic features and CAZomes

To assess the genomic diversity in the genus *Pycnoporus*, the genome of *P. coccineus* BRFM 310 (herein named Pycco), *P. sanguineus* BRFM 1264 (Pycsa) and *P. puniceus* BRFM 1868 (Pycpun) were sequenced and compared with that of *P. cinnabarinus* BRFM 137 (herein named Pycci).[9] The genome size of *P. coccineus*, *P. sanguineus* and *P. puniceus*, ranging from 30 to 36 Mb were in line with that of *P. cinnabarinus* (33.67 Mb) and WR relatives (Table 1; Supplementary Fig. S1 and Table S1). We observed low amounts of repeat sequences (1.8–12.3%) and the absence of major rearrangements in the genomes (Supplementary Figs S2 and S3). The genes coding for mating type, Class II peroxidases, CAZymes, peptidases, GSTs, hydrophobins, proteins from the secretory pathway, the glycosylation pathway, the carbon catabolism pathway and SSPs were inspected by expert annotation (Supplementary Tables S2–S15 and Figs S4–S13). We observed a high proportion of conserved protein-coding genes across the genomes (82.3% of the *P. cinnabarinus* protein-coding genes) and a low proportion of species-specific genes (4–5%; Supplementary Fig. S14). Inspection of mating type genes showed high sequence identity between the alleles of the four *Pycnoporus* species (Supplementary Table S4). In particular, *P. coccineus* BRFM 310 and *P. sanguineus* BRFM 1264 alleles were much more similar to each other than to that of the two other *Pycnoporus* species, in support of the notion that these two species are closely related.[15,50]

The CAZyme gene repertoires (CAZome) in the three newly sequenced genomes were similar to that of *P. cinnabarinus* (mean 436 CAZymes classified into 108 CAZy families). The gene counts for FCW-degrading enzymes were similar to that of other Polyporales fungi (Supplementary Fig. S15). As a common feature of WR fungi,[4,51] *Pycnoporus* genomes contained a large number of genes coding for PCW-active CAZymes when compared with brown-rot fungi, which use Fenton chemistry in combination with a limited number of CAZymes for PCW breakdown (Fig. 1, Supplementary Figs S16–S18). The genomes were rich in lytic polysaccharide monooxygenases (LPMOs) active on crystalline cellulose and β-(1,4)-linked hemicellulose polysaccharides (CAZy family AA9; 13–17 genes). The Class II PODs (in total 9–11 genes) were identified as manganese peroxidases (MnP), versatile peroxidases (VP) and lignin peroxidases (LiP; Supplementary Fig. S19 and Table S5). We observed a high number of predicted secreted oxidoreductases that could act as AA enzymes for the oxidative degradation of PCWs. Among them, GMC-oxidoreductases from CAZy subfamily AA3_2 (20–22 genes) have pivotal roles in PCW degradation. Secreted AA3_2 are involved in the generation of H_2O_2, a co-substrate for class II peroxidases and LPMOs. AA3_2 also contribute to the oxidation of saccharides and to the redox cycling of aromatic alcohols and quinones. In addition, we identified three AA5_1 glyoxal oxidase genes in each of the genomes, which encode copper radical oxidases involved in extracellular H_2O_2 production (Supplementary Figs S9–S13 and Table S6).

3.2. *Pycnoporus* species show diverse responses to lignocellulosic substrates

To assess the functional diversity within the genus *Pycnoporus*, we compared the ability of the phylogenetically most closely related species; *P. cinnabarinus*, *P. coccineus* and *P. sanguineus* to grow on a variety of plant-derived carbon sources on agar plate assays. In these conditions, we observed differences in fungal growth on complex lignocellulosic substrates (Fig. 2a and Supplementary Fig. S20). We next analysed the early response of the three species to cellulose, WS and woody substrates in agitated liquid culture media and compared their transcriptomes and the secreted proteins collected from the media. Maltose was used as a control, as it was shown not to induce carbon catabolic repression in ascomycetes.[52] AVI was used as a cellulose-enriched substrate, and WS, pine and Asp were used as representatives of Gramineae, softwood and hardwood, respectively. At day 3, most cultures had initiated growth and consumed all maltose initially present in the medium. This time-point was therefore selected to analyse the early response of each species to the substrates (Fig. 3; Supplementary Figs S21 and S22).

The global transcriptome responses varied among the species. Especially, *P. coccineus* showed the highest proportion of regulated genes with up to 7.6% of the genes up-regulated on pine (fold change in transcript abundance $\geq$4; Fig. 2b). Similarity between the transcriptomes of the three species was assessed by analysing the differential expression of one-to-one orthologous genes (co-orthologs) after 3-day growth on cellulose, WS or Asp when compared with maltose. To identify co-orthologs, in silico-deduced proteomes of the three species were clustered into 13,836 orthogroups using OrthoFinder, of which 6,524 represented co-orthologs. Similar

Table 1. Features of *P. coccineus* BRFM 310, *P. puniceus* BRFM 1868 and *P. sanguineus* BRFM 1264 genome assemblies and annotations

	P. cinnabarinus	*P. coccineus*	*P. puniceus*	*P. sanguineus*
Genome size (Mbp)	33.67	32.76	30.26	36.04
Number of contigs	2,036	469	105	2,046
Number of scaffolds	784	222	105	657
Scaffold N50	54	20	12	35
Scaffold L50 (Mbp)	0.17	0.47	0.79	0.32
Transposable element coverage (%)	8.15	1.8	12.33	4.91
Transposable element coverage (Mbp)	2.74	0.59	3.73	1.77
Number of predicted proteins	10,442	12,690	10,050	14,165
BUSCO complete protein sequences	1,268	1,321	1,309	1,293
BUSCO fragmented protein sequences	32	7	14	21
BUSCO missing protein sequences	35	7	12	21

The reliability of gene structural annotations was assessed using universal single-copy orthologs (BUSCO). The genome of *P. cinnabarinus* BRFM 137[9] is indicated for comparison.

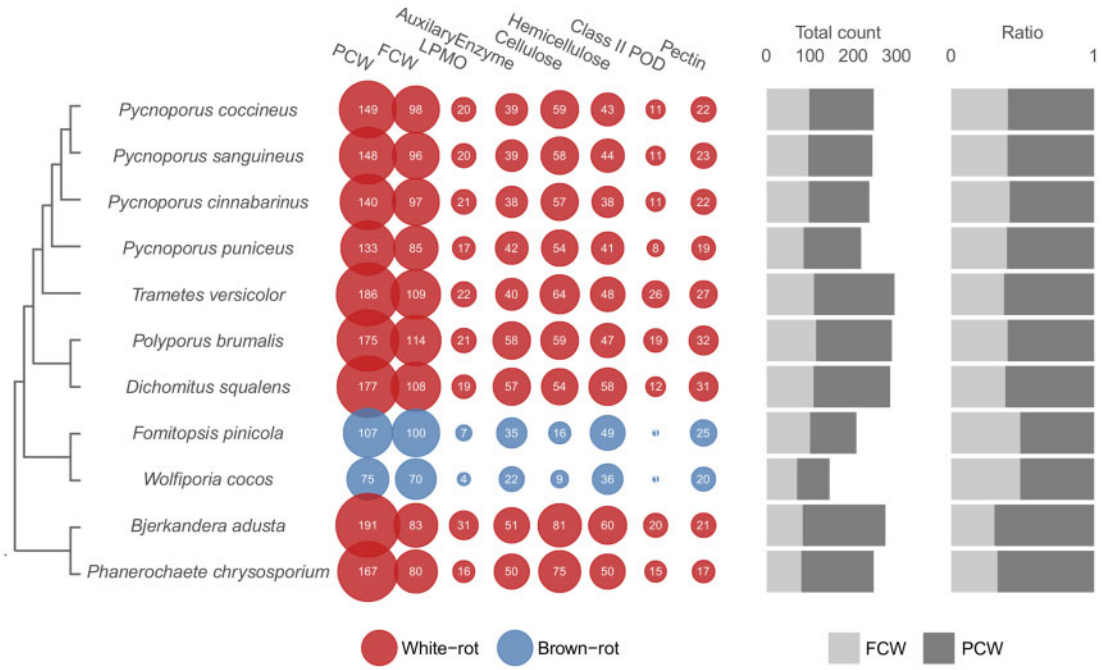


Figure 1. Gene counts for CAZyme domains of PCW and FCW-degrading enzymes. The bar plots show the total count of genes including PCW and FCW-degrading enzymes (left); and the ratio of PCW to FCW-degrading enzymes (right). The counts for AA enzymes that could contribute to PCW degradation include AA1_1 laccases and predicted secreted AA3, AA4 and AA5. The counts for PCW-active LPMOs include AA9, AA13, AA14 and AA16. Enzymes active on FCWs, cellulose, hemicellulose or pectin were classified according to Supplementary Figs S15–S18.

transcript regulations were frequent between co-orthologs of two species. Surprisingly, however, we observed poor conservation of transcript regulation of co-orthologs across the three species, including for genes with high transcription induction or repression on particular carbon sources (Fig. 4a). We examined the transcript read counts of CAZyme-coding genes. Transcriptome profiles were more similar within the species cultured under the different conditions than between the species cultured on the same substrates (Supplementary Fig. S23). Also, we observed in each species that approximately half of the regulated CAZymes (41–51%) were up-regulated in response to cellulose, WS and Asp, highlighting the presence of core regulations to (ligno)cellulosic substrates with diverse compositions (Supplementary Fig. S24).

3.3. Conserved gene regulations in response to lignocellulosic substrates

In search for conserved enzymatic mechanisms involved in the initiation of lignocellulose breakdown, we analysed the gene sets sharing similar transcription profiles at the onset of the response to cellulose, WS or Asp. We focused on genes coding for proteins typically found in fungal exo-proteomes, i.e. CAZymes, peptidases, SSPs and hydrophobins.[53] A total of 2,227 manually curated genes from the genomes of *P. cinnabarinus*, *P. coccineus* and *P. sanguineus* were analysed. In order to combine all RNA-seq data in a single cross-species analysis, transcript read counts were normalized using the DESeq2 package,[43] the normalized read counts were log2 transformed and subjected to removal of batch effects and to quantile normalization. We checked the impact of each normalization step on the

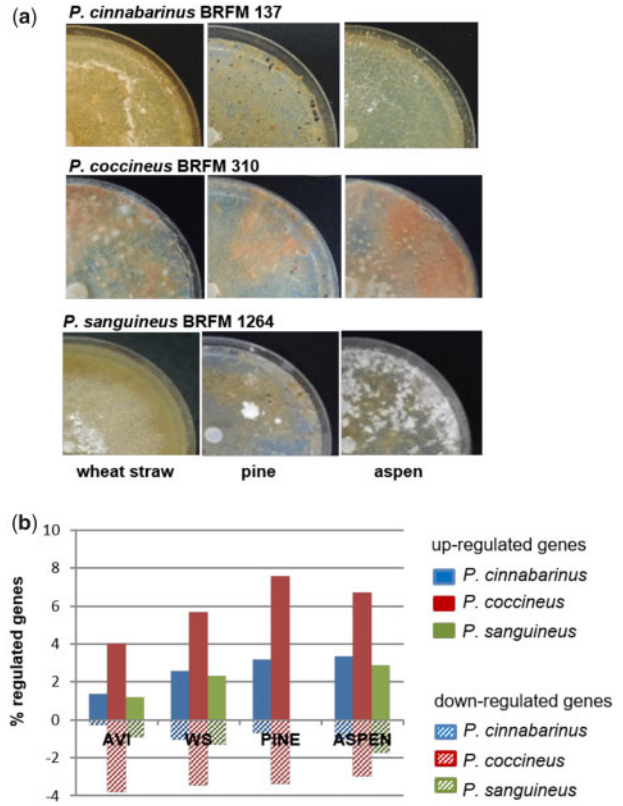


Figure 2. Phenotype polymorphism across three *Pycnoporus* species. (**a**) Agar plates after 6 weeks cultivation on ground WS, pine or Asp. *P. cinnabarinus* BRFM 137 did not develop mycelium on pine and Asp. The white dots formed by *P. sanguineus* on pine and Asp are arthrospores indicating that the fungus stopped growing to form dormant structures. (**b**) Percentage of regulated genes after 3-day growth in liquid cultures on AVI, WS, pine or Asp when compared with maltose (|fold change| ≥ 4) in the three *Pycnoporus* species. No RNASeq data was available for *P. sanguineus* grown on pine.

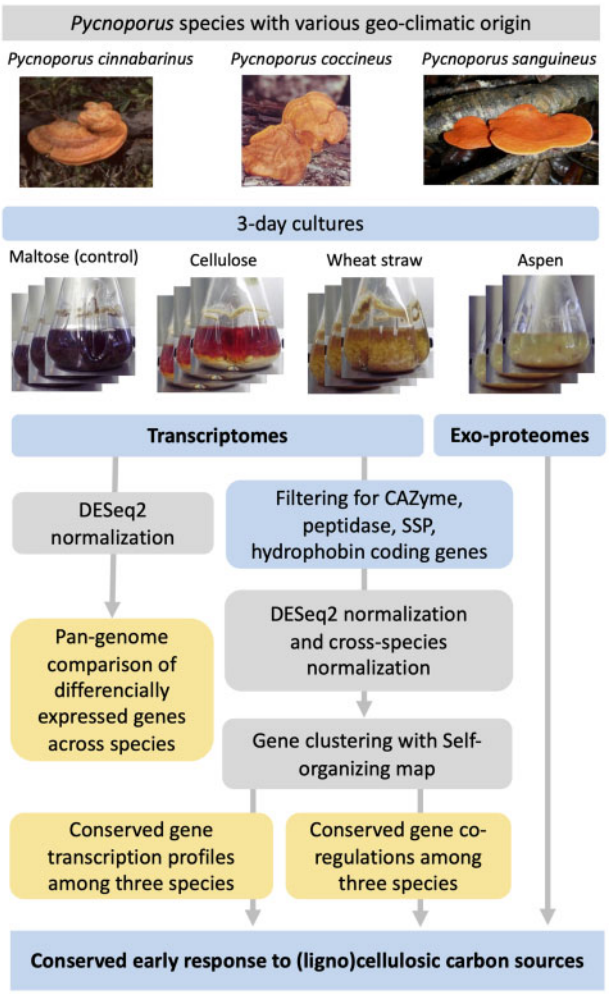


Figure 3. Cross-species comparison of the early response of three *Pycnoporus* species to lignocellulosic carbon sources.

distribution of the data (Supplementary Figs S25–S28). Inspecting the one-to-one co-orthologs from this set of genes (405 orthology groups), we observed low conservation of the normalized transcript read counts across the species, except for growth on maltose and cellulose, and for *P. coccineus* and *P. sanguineus* grown on Asp, indicating that complex lignocellulosic substrates induced more diverse responses across the species (Fig. 4b).

We grouped genes with similar transcript profiles into nodes (clusters of co-regulated genes) using the SOM unsupervised learning method. SOM is a data-driven clustering method constructing a topographic organization of nodes in which neighbouring nodes share similar transcriptome patterns, and thereby alleviates the requirement for arbitrary clustering thresholds. SOM allows the co-localization on the SOM map of genes with similar transcript levels. We used SOM on the normalized read counts from the three species in order to identify both the co-regulated genes within each species and the genes with similar transcript patterns across the species. We produced 72 nodes containing on average 31 genes per node (Fig. 5a and b). Exo-proteome data (Supplementary Tables S16–S18) were combined to the SOM map to integrate gene transcription and protein secretion information.

We first examined the transcriptional response to cellulose. We identified 19 nodes containing 250 genes highly transcribed (mean normalized log2 read count ≥12) or up-regulated (mean fold change ≥4 compared with maltose; Supplementary Table S19). Inspecting the gene content for these nodes (e.g. Fig. 5c), we looked for shared gene differential expression across the three species. In the three species, we found up-regulation for CAZyme-coding genes involved in β-(1,4)-glucan linkage breakdown including endoglucanases from family 5 subfamily 5 (GH5_5), GH45 and GH131, cellobiohydrolases (GH6, GH7), cellobiose dehydrogenases (CDH) and AA9 LPMOs (Fig. 6a). Also, there was a wide panel of enzymes active on hemicelluloses from families GH10, GH74, GH5_7, GH12, GH115, CE1, CE15, CE16, enzymes active on pectin CE8, GH28 and enzymes with promiscuous activities on glycosidic bounds GH1 and GH3. Among the 20 CAZyme families identified, 14 were represented by co-orthologous genes from the 3 species, showing conservation of their regulation in response to cellulose across the genus. Globally, 50% of the enzymes encoded by the up-regulated genes were detected in the culture medium (Fig. 6a), strengthening a role for these enzymes as key players in lignocellulose breakdown by *Pycnoporus* fungi.

To identify the genes specifically regulated for the complex lignocellulosic substrates, we selected genes up-regulated on WS or Asp, not on cellulose, when compared with maltose. There were 13 nodes including 225 genes that met at least one of the following criteria; (i) average normalized log2 read counts >12 on WS or Asp and <12 on

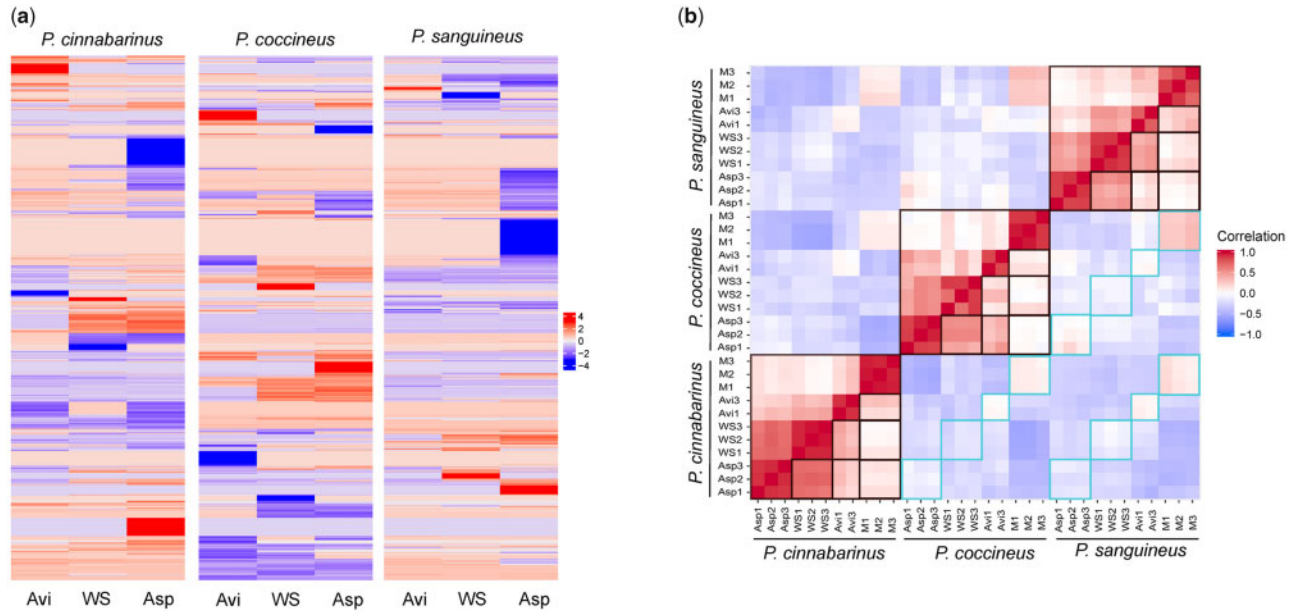


Figure 4. Global transcriptome similarity between co-orthologous genes from *P. cinnabarinus*, *P. coccineus* and *P. sanguineus*. (**a**) Heatmap of changes in transcript read counts (log2 fold change) after 3-day growth on each carbon source when compared with maltose for 6,524 groups of 1-to-1 co-ortholog genes. (**b**) Pearson correlation coefficient for the normalized transcript read counts in each growth condition for the 405 1-to-1 co-ortholog CAZyme, peptidase, SSP and hydrophobin genes identified in the genomes. The comparisons of the response of each species to various substrates are highlighted in black boxes. Cross-species comparisons on a same substrate are highlighted in blue boxes. M: maltose; AVI: Avicel; WS: wheat straw; Asp: aspen.

maltose and cellulose, or (ii) mean fold change ≥4 on WS or Asp, and ≤4 on cellulose when compared with maltose. From these nodes, we identified the gene families with homologs among the three species. These included genes coding for enzymes that target hemicelluloses (CE16 and CE4 acetylesterases, GH2 *β*-mannosidases or *β*-glucuronidases, GH16 xyloglucan hydrolase, GH27 *α*-galactosidases, GH30 *β*-glucosidase/*β*-xylosidase, GH51 *α*-L-arabinofuranosidases and GH31 *α*-xylosidases/*α*-glucosidases), enzymes that target pectin (GH28 polygalacturonases, PL14_4 *β*-1,4-glucuronan lyases), AA1_1 laccases and AA3_2 GMC-oxidoreductases (Fig. 6b, Supplementary Table S20). Genes coding for glycoside hydrolases active on the FCW (GH18 chitinases, GH76 *α*-mannanases) and peptidases were also specifically up-regulated in the three species.

The genomes of the three species encoded the three types of class II peroxidases involved in lignin breakdown; MnP, LiP and VP (Supplementary Table S5). Orthologous MnP and LiP genes (defined by protein sequence phylogeny; Supplementary Fig. S6) did not show conserved regulation across the species in response to lignocellulosic substrates (Fig. 7). The highest induction factors were found in *P. coccineus* and a MnP (protein ID #1468611)- and a LiP (#1431101)-coding gene reached 800- and 1,500-fold induction on Asp, respectively. For the three species, we observed no up-regulation of the VP-coding genes.

3.4. Co-regulated genes indicative of enzymatic synergies

We investigated potential enzymatic synergies conserved in the three *Pycnoporus* species through co-regulated gene transcription with co-secreted corresponding proteins. In each species, one single CDH gene was present in the genome, which shared a similar transcription pattern with AA9 LPMO-coding genes (nodes 39, 57 and 58; Figs 5c and 8). These AA9 LPMO and CDH genes were up-regulated on cellulose and the corresponding proteins were identified in the exoproteomes. AA9 LPMOs oxidatively cleave glycosidic chains at the cellulose surface thereby creating new substrate-binding sites for hydrolytic cellulases.[54] CDHs are composed of a flavin adenine dinucleotide-binding dehydrogenase domain (AA3_1) connected via a flexible linker to a haem *b*-binding cytochrome domain[55] (Cyt*b*; AA8). *In vitro*, CDHs behave as electron donors for AA9 LPMOs and boost LPMO activity (reviewed in Berrin *et al.*[56]). Our results suggest CDH as a biologically relevant electron donor for AA9 LPMOs.

As indicated above, genes coding for predicted class II peroxidases (AA2s) were specifically up-regulated on WS or Asp, but not on cellulose (Fig. 5b). Node 40, 41 and 32 contained three *mnp* genes from *P. coccineus* co-regulated with the three predicted glyoxal oxidase genes identified in this genome, suggesting that glyoxal oxidases might provide these MnPs with the H_2O_2 required for their activity (Fig. 6). In contrast in *P. cinnabarinus* and *P. sanguineus*, we did not observe any co-regulations at the transcript level between class II peroxidases and glyoxal oxidases.

3.5. Differentially regulated genes indicative of detoxification

In natural environments, lignocellulose degradation may lead to the release of toxic degradation products and extractives (e.g. phenols, terpenoids and tannins).[57,58] We investigated if the regulation of genes involved in detoxification was conserved across the three *Pycnoporus* species in our growth conditions. For this purpose, we examined genes coding for putative GST in the cross-species SOM analysis.

We identified a GST gene from class GTT2 in *P. sanguineus* that was clustered in node 31 filtered for high transcript level on WS and Asp, and one GTT2 gene from *P. cinnabarinus* that was clustered in node 22 filtered for up-regulation on WS (Supplementary Table S20). One isoform of this class from *Phanerochaete chrysosporium*

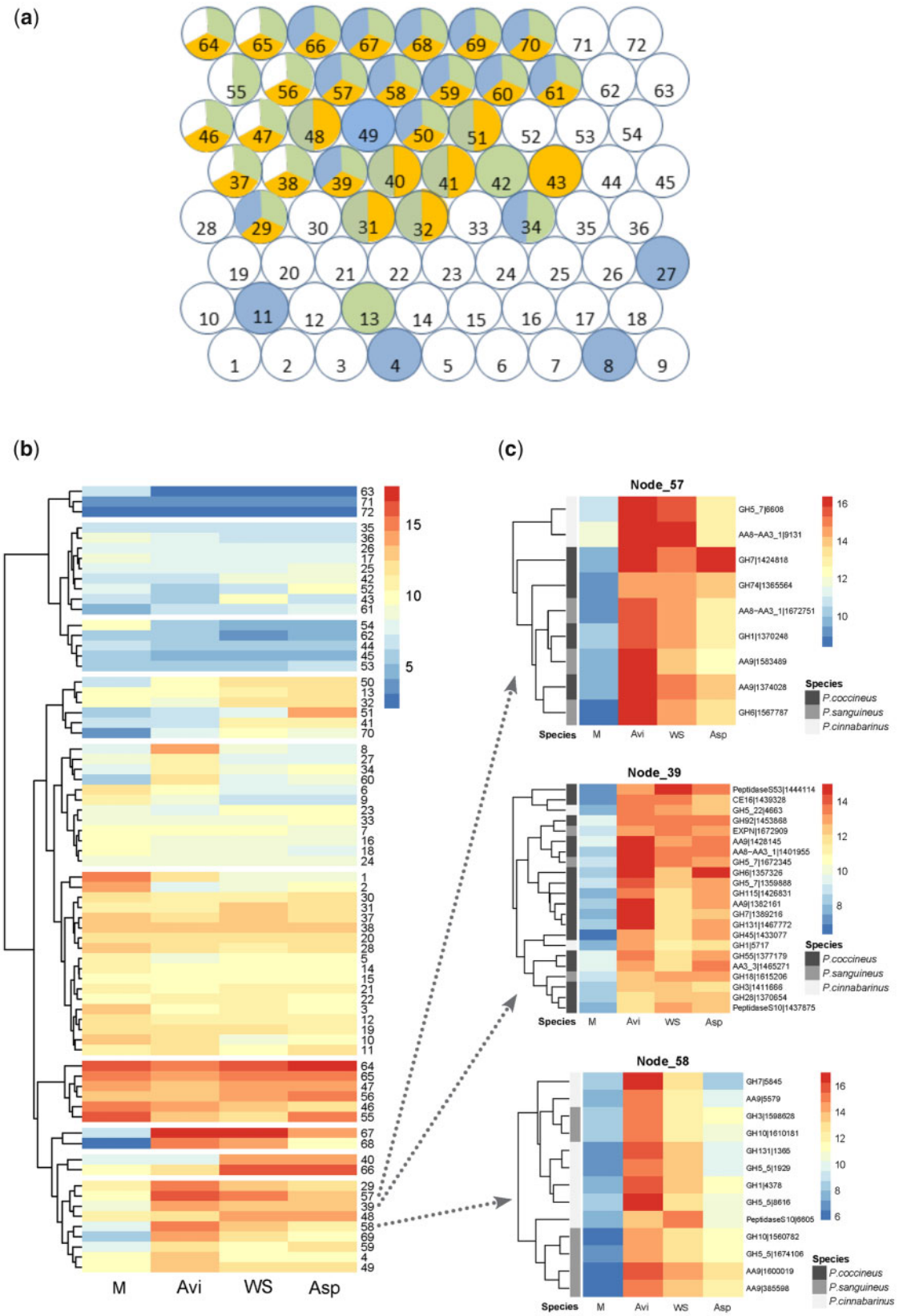


Figure 5. Clustering of genes coding for predicted CAZymes, peptidases, SSPs and hydrophobins in three *Pycnoporus* species according to their transcription profile on maltose (M), AVI, WS and Asp. (**a**) SOM clustering resulted in 72 nodes with average 31 genes per node. Nodes containing genes highly transcribed or up-regulated on cellulose (blue), Asp (green) or WS (orange) when compared with maltose are highlighted.(**b**) Hierarchical clustering of the nodes according to the averaged normalized transcript read counts on each carbon source. (**c**) Gene content and transcript profiles of nodes 57, 39 and 58. AA8-AA3-1: cellobiose dehydrogenase.

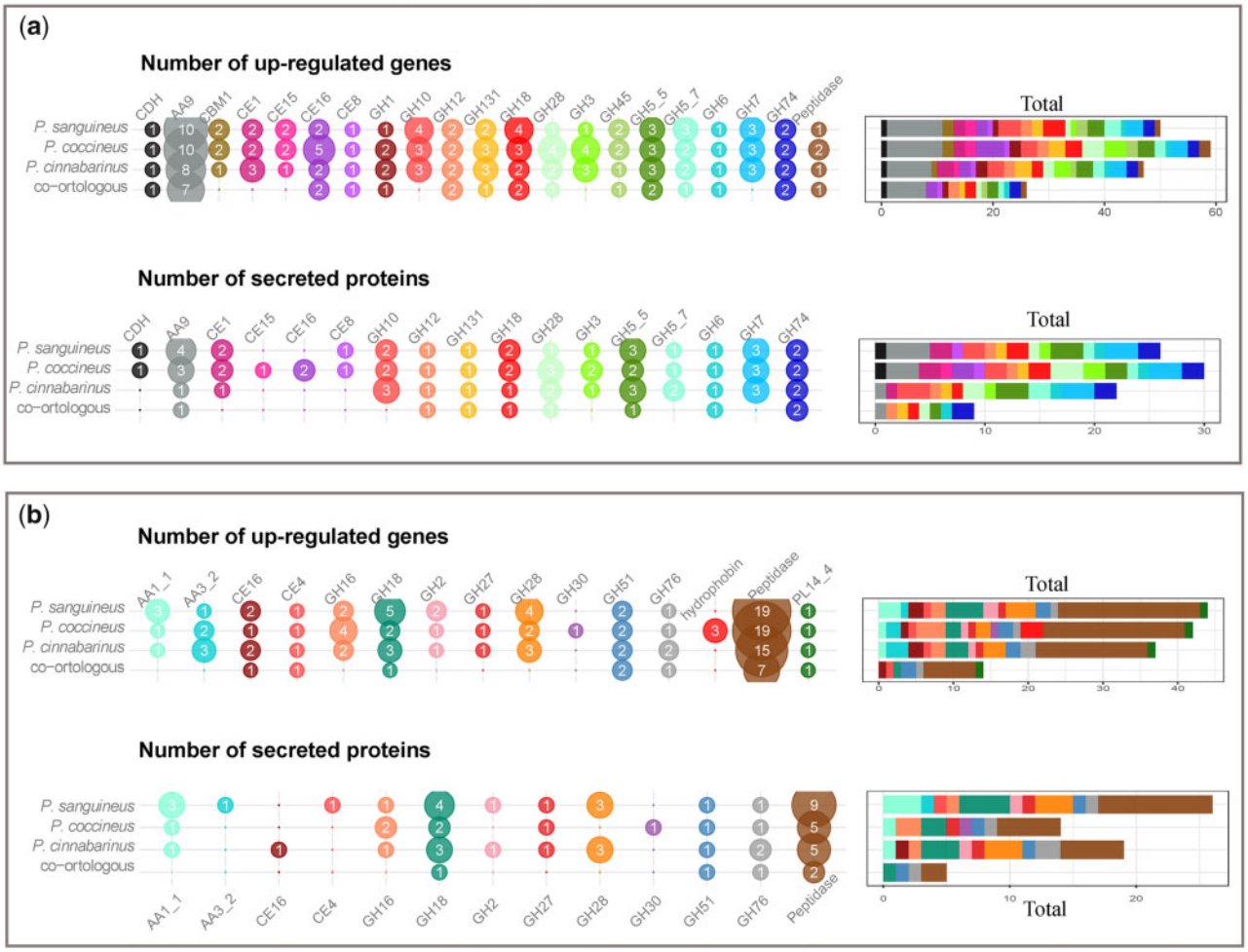


Figure 6. Shared expression regulation of CAZyme genes across the three *Pycnoporus* species. (**a**) Numbers of genes with shared differential expression on cellulose and numbers of proteins secreted during growth on cellulose. (**b**) Numbers of genes with shared specific differential expression on lignocellulosic substrates, not on cellulose. The numbers of orthologous groups of genes with conserved transcription regulation and secretion across the three species are indicated.

has been functionally characterized and has been shown to reduce high-molecular weight peroxides.[59,60] In contrast, a total of four GST genes of *P. coccineus* were regulated on the tested substrates; two GST class A (FuA) genes were up-regulated on all the tested substrates (Supplementary Table S19) whereas one S-glutathionyl hydroquinone reductase and one GST Omega were specifically up-regulated in response to WS or Asp (Supplementary Table S20). For the case of *P. chrysosporium*, these classes of GSTs have potential roles in the detoxification of wood-derived molecules acting as ligandins for wood extractives or as catalysts for deglutathionylation of various substrates, including hydroquinone conjugates and terpenes.[36,59,61] These findings suggest that different GST-related detoxification systems may exist in *P. coccineus*.

Cytochrome P450 (CYP450) monooxygenases are commonly involved in the first step of eliminating toxic molecules including molecules released from lignocellulose. CYP450 can initiate the modification of these molecules via hydroxylation, epoxidation or monooxygenation.[62,63] The genomes of *P. cinnabarinus*, *P. coccineus* and *P. sanguineus* contain 107, 132 and 113 predicted CYP450, respectively. Inspection of the transcript profiles for these genes showed that *P. coccineus* had the highest number of up-regulated CYP450 genes in response to the tested substrates (17 genes) compared with *P. cinnabarinus* and *P. sanguineus* (14 and 6 up-regulated genes, respectively). Some CYP450 genes were commonly up-regulated in at least two species, which were from families CYP63, CYP5035, CYP5139, CYP5144 and CYP5150 According to the Fungal cytochrome P450 database[64] (Supplementary Table S13). Of these, CYP63, CYP5139 and CYP5144 were shown to be active on multiple xenobiotic compounds such as polycyclic aromatic hydrocarbons, alkylphenols and alkanes.[63,65,66] CYP5035s oxidize plant chemicals (i.e. resin and flavonoids), and the number of genes for CYP5035 and CYP5150 expanded in basidiomycetes that grow on wood or litter.[63] Our results suggest that these CYP450 families could be differently utilized by the species for detoxification of molecules released from the lignocellulosic substrates.

4. Discussion

4.1. Phenotype diversity outreaches diversity of the predicted proteome across the genus *Pycnoporus*

The four *Pycnoporus* species studied here are found in different geoclimatic areas[15,16] covering the Northern hemisphere (*P.*

species		Prot ID	normalized log2 read count				M Avi WS Asp
			M	Avi	WS	Asp	
P. cinnabarinus	LiP	7499	7,23	6,46	7,20	8,68	
P. coccineus	LiP	1431101	6,04	5,35	9,27	15,59	
P. sanguineus	LiP	1603832	7,06	6,87	7,30	8,39	
P. cinnabarinus	LiP	7860	6,56	5,87	5,48	5,47	
P. coccineus	LiP	859168	7,23	6,30	6,54	6,66	
P. sanguineus	LiP	1748658	9,03	7,59	7,51	7,59	
P. cinnabarinus	LiP	7500	10,50	9,63	9,18	9,93	
P. coccineus	LiP	1403742	4,42	3,91	4,59	3,94	
P. sanguineus	LiP	1676767	7,70	6,71	5,72	6,43	
P. coccineus	VP	1469331	5,27	4,54	5,26	5,35	
P. sanguineus	VP	1607934	5,48	5,40	5,54	4,82	
P. cinnabarinus	LiP	7501	6,61	5,77	5,80	5,37	
P. coccineus	LiP	779035	5,73	3,96	5,88	5,16	
P. sanguineus	LiP	1676766	5,40	4,64	4,69	4,99	
P. cinnabarinus	VP	6829	5,72	5,49	6,26	6,23	
P. coccineus	VP	1468768	6,02	4,02	5,37	6,01	
P. sanguineus	VP	1587729	8,43	7,43	7,55	8,25	
P. coccineus	MnP	1369658	7,40	6,58	6,42	7,26	
P. sanguineus	MnP	1672299	6,35	4,98	5,68	5,87	
P. cinnabarinus	MnP	737	9,52	6,27	6,19	12,26	
P. coccineus	MnP	1468611	6,51	6,46	14,34	15,39	
P. sanguineus	MnP	1603419	7,96	5,95	7,45	9,72	
P. cinnabarinus	MnP	7498	9,32	8,07	7,34	8,23	
P. coccineus	MnP	1464049	7,40	5,78	10,46	13,35	
P. sanguineus	MnP	1651739	11,29	3,37	4,53	4,04	
P. cinnabarinus	MnP	480	4,18	4,65	3,59	8,10	
P. coccineus	MnP	1436321	7,27	7,10	13,86	12,92	
P. sanguineus	MnP	35461	7,45	5,98	6,85	8,85	
P. cinnabarinus	VP	6481	9,54	8,69	9,26	8,60	
P. coccineus	VP	1438352	12,29	9,50	8,44	7,75	
P. sanguineus	VP	1679991	12,62	7,52	7,03	7,19	

Figure 7. Regulation of Class II peroxidase genes in response to the substrates. Orthologous genes are grouped for comparison of their transcription profiles.

cinnabarinus), countries bordering the Indian and Pacific Oceans (*P. coccineus*), the tropics and subtropics of both hemispheres (*P. sanguineus*) and paleotropical areas (*P. puniceus*), with overlap limited to Japan and Eastern China. Accordingly, the parental strains of the monokaryons studied here were collected in Europe, China, South America and Indonesia, respectively. Here, we show that the genomes sequenced from geographically distant strains did not show any major rearrangements in structure and protein-coding gene composition. Our findings suggest that geographical isolation of the species the genus *Pycnoporus* has not yet lead to any noticeable genome level diversity. In addition, the high sequence identity between mating type genes of the four species, particularly between *P. coccineus* and *P. sanguineus* suggests that the speciation events are recent in this genus. In the genus *Laetiporus*, another group of wood-decay fungi from the order Polyporales that originated about 20 mya,[67] several regional speciation events have been estimated between 9.88 and 2.89 mya. Some of these speciation events have been related to vicariance that emerged due to tectonic activity and the disconnection of the continental plates or from wind and ocean current dispersion of basidiospores.[67] It is currently estimated that the genus

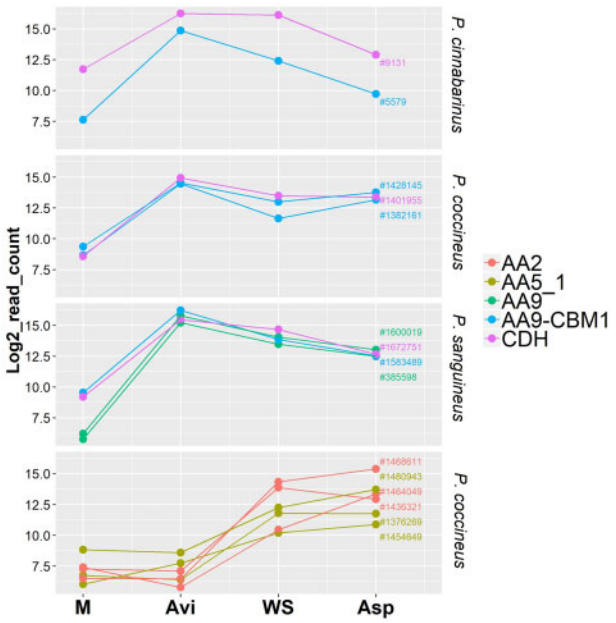


Figure 8. Conserved co-regulations of CDH and AA9 LPMO genes in *P. cinnabarinus, P. coccineus* and *P. sanguineus* and co-regulations of MnP- and GLOX-genes in *P. coccineus*. Transcript levels are expressed as log2-transformed read counts after 3-day-growth on maltose (M), AVI, WS or Asp. CBM: carbohydrate-binding module.

Pycnoporus originated about 40–50 mya (Nagy LG, personal communication). Further molecular studies are needed to understand the dispersal routes and the role of regional speciation in the evolutionary history of *Pycnoporus*.

Despite conserved CAZomes, the *P. cinnabarinus*, *P. coccineus* and *P. sanguineus* strains analysed here showed different abilities to grow on (ligno)cellulosic substrates. This finding demonstrates that gene repertoires obtained from genome sequencing are not sufficient to predict the ability of a strain to grow on recalcitrant raw biomass. The comparison of the early responses of the fungi to the substrates showed that gene transcription profiles better correlated with the strains rather than with the substrates. A similar trend was observed in the WR fungus *Pleurotus ostreatus*, between monokaryotic strains issued from a same parental dikaryotic strain. In *P. ostreatus*, these differences in gene regulation have been partly attributed to the presence of transposable elements near the differentially regulated genes.[68] A closer analysis of transposable element localization or methylation status of each monokaryon haplotype would help understanding the transcription regulation we observed.

4.2. Conserved sets of enzymes produced at the onset of wood-decay activity

As a conserved response to cellulose across three *Pycnoporus* species, we identified here a set of CAZyme genes with shared transcriptional regulations. For a sub-set of these genes, the corresponding enzymes were commonly found in the secretomes of the three species at the analysed time point, further strengthening a role for these enzymes as key players in lignocellulose breakdown by these fungi. These included genes coding for enzymes active on β-1,4-glucan linkages (GH5_5 and GH131 endo-β-1,4-glucanases, GH6 and GH7 cellobiohydrolases and AA9 LPMOs) and genes targeting the hemicellulose backbone (GH10 endoxylanases, GH12 and GH74 xyloglucan hydrolases, GH5_7 β-mannosidases) or branched groups (CE1 feruloyl esterases) and enzymes active on pectin (GH28 polygalacturonases). In the presence of lignocellulosic substrates, additional genes were commonly induced and the corresponding CAZymes secreted, that target hemicelluloses (GH16 xyloglucan hydrolase, GH27 β-galactosidases, GH30 β-glucosidases/β-xylosidases, GH51 α-L-arabinofuranosidases), peptidases and laccases (AA1_1).

4.3. AA9 LPMOs and their enzymatic partners

AA9 LPMOs are known as key players in oxidative cellulose depolymerization by wood-decay fungi.[56] Expansions in AA9 gene numbers is a common feature of the genomes of WR fungi,[69] but whether members of the AA9 gene family play different roles during wood decay has not been elucidated. We identified 17, 16 and 11 genes coding for AA9 LPMOs in *P. cinnabarinus*, *P. coccineus* and *P. sanguineus*, from which 11 (64%), 13 (81%) and 10 (90%), respectively, were up-regulated during the early response to cellulose-containing substrates. In each strain, several AA9 LPMO genes shared similar transcription profiles, and several AA9 LPMOs were concomitantly secreted, indicating that several AA9 LPMOs co-secreted by the fungi might act simultaneously on the substrate.

The identification of enzymatic partners for LPMOs is a very active field of research.[70–73] Looking at conserved co-regulations across the three strains, we identified CDHs as invariably co-regulated with at least one AA9 LPMO gene. CDHs have been shown to reduce O_2 and generate H_2O_2 which in turn can fuel, via the Fenton reaction, the production of hydroxyl radicals that disrupt lignocellulose polymers.[74] More recently, CDHs have been shown to activate AA9 LPMOs *in vitro* through electron transfer[75] and it has been proposed that CDHs could also activate AA9 LPMOs through the generation of H_2O_2, a co-substrate for the enzyme.[76] Our findings support the hypothesis of synergy between AA9 LPMOs and CDHs in vivo as a conserved mechanism for cellulose degradation in the genus *Pycnoporus* spp.

4.4. Co-secretion of extracellular H_2O_2-generating and -consuming enzymes

Extracellular H_2O_2 is a central factor in oxidative lignocellulose breakdown. It was proposed that WR fungi could avoid oxidative damages due to H_2O_2 accumulation in the vicinity of the hyphae by the co-secretion of H_2O_2-generating and H_2O_2-consuming enzymatic partners.[72] Some secreted GMC-oxidoreductases such as glucose dehydrogenases and aryl-alcohol quinone oxidoreductases (AA3_2), which generate H_2O_2, are able to prime AA9 LPMO activity *in vitro*.[70,73] Similarly, glyoxal oxidases can generate H_2O_2 and prime MnP and LiP activity *in vitro*.[77] A biological relevance for these synergetic activities was previously suggested from the co-occurrence in fungal secretomes of class II peroxidases with GMC-oxidoreductases[10,51,78–82] or glyoxal oxidases.[10,51,79,81,83–85] Here, we confirmed the co-occurrence in the secretomes of GMC-oxidoreductases and glyoxal oxidases with class II peroxidases and AA9 LPMOs as a common feature of three *Pycnoporus* species. However, we detected no evidence for co-regulation of the corresponding genes at the transcript level, except between three MnPs and two glyoxal oxidase encoding genes in *P. coccineus*. We hypothesize that independent tight regulations of gene expression for H_2O_2-generating and H_2O_2-consuming enzymes at the transcription level could allow a more flexible system to rapidly adapt extracellular reactive oxygen species concentrations and avoid oxidative damage to the hyphae.

5. Conclusion

Although WR fungi are well-studied for the diversity of enzymatic systems for lignocellulose breakdown, their intra-genus diversity has not been extensively explored so far. Here, we investigated the functional diversity within the genus *Pycnoporus*, a group of fungi from the Trametes clade renowned for the ecological significance and a great potential for novel biocatalysts. We introduced a new method for cross-species comparisons of transcriptomes, using the genome sequences of the four species described in the genus and expert gene functional annotations. The transcriptomic aspect of observations was validated with experimentally identified secreted proteins, supporting our hypotheses. We identified a conserved set of genes responsive to lignocellulosic substrates that outlines key enzymatic mechanics for wood decomposition activity in these fungi. We observed that the mechanisms involved in oxidative lignin breakdown were more diverse than those in polysaccharide breakdown and the fungal enzymes detoxifying lignocellulose-released products were different among the species.

Acknowledgements

We are grateful to Valentin Loux from the INRA MIGALE bioinformatics platform (http://migale.jouy.inra.fr) and to Estelle Bonnin (INRA, France) for substrate composition analysis. Magali Lescot (MIO, France) and Oswaldo Forey (INRA) provided fruitful advises for comparative genomics and graphical output, respectively. Miaomiao Zhou (Westerdijk Fungal Biodiversity Institute) provided support for gene expert annotation.

Accession numbers

NCSW00000000, VOCM00000000, VICQ00000000, GSE82486.

Funding

This work was supported by The French National Agency for Research (ANR-14-CE06-0020). H.H. was granted by the Institut national de recherche pour l'agriculture, l'alimentation et l'environnement and Région SUD Provence-Alpes-Côte d'Azur. The work conducted by the U.S. Department of Energy (DOE) Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The work by the Consejo Superior de Investigaciones Científicas was supported by the Ministerio de Ciencias, Innovación y Universidades de España (BIO2017-86559-R).

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at DNARES online.

References

1. Rajala, T., Peltoniemi, M., Pennanen, T. and Mäkipä, R. 2012, Fungal community dynamics in relation to substrate quality of decaying Norway spruce (*Picea abies* [L.] Karst.) logs in boreal forests, *FEMS Microbiol. Ecol.*, **81**, 494–505.
2. Valentín, L., Rajala, T., Peltoniemi, M., Heinonsalo, J., Pennanen, T. and Mäkipää, R. 2014, Loss of diversity in wood-inhabiting fungal communities affects decomposition activity in Norway spruce wood, *Front. Microbiol.*, **5**, 230.
3. Blanchette, R.A. 1991, Delignification by wood-decay fungi, *Annu. Rev. Phytopathol.*, **29**, 381–403.
4. Riley, R., Salamov, A.A., Brown, D.W., et al. 2014, Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi, *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 9923–8.
5. Couturier, M., Ladevèze, S., Sulzenbacher, G., et al. 2018, Lytic xylan oxidases from wood-decay fungi unlock biomass degradation, *Nat. Chem. Biol.*, **14**, 306–10.
6. Fernandez-Fueyo, E., Ruiz-Dueñas, F.J., Ferreira, P., et al. 2012, Comparative genomics of *Ceriporiopsis subvermispora* and *Phanerochaete chrysosporium* provide insight into selective ligninolysis, *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 5458–63.
7. Kuuskeri, J., Mäkelä, M.R., Isotalo, J., Oksanen, I. and Lundell, T. 2015, Lignocellulose-converting enzyme activity profiles correlate with molecular systematics and phylogeny grouping in the incoherent genus *Phlebia* (*Polyporales*, *Basidiomycota*), *BMC Microbiol.*, **15**, 217.
8. Hastrup, A.C.S., Howell, C., Larsen, F.H., Sathitsuksanoh, N., Goodell, B. and Jellison, J. 2012, Differences in crystalline cellulose modification due to degradation by brown and white rot fungi, *Fungal Biol.*, **116**, 1052–63.
9. Levasseur, A., Lomascolo, A., Chabrol, O., et al. 2014, The genome of the white-rot fungus *Pycnoporus cinnabarinus*: a basidiomycete model with a versatile arsenal for lignocellulosic biomass breakdown, *BMC Genomics*, **15**, 486.
10. Miyauchi, S., Rancon, A., Drula, E., et al. 2018, Integrative visual omics of the white-rot fungus *Polyporus brumalis* exposes the biotechnological potential of its oxidative enzymes for delignifying raw plant biomass, *Biotechnol. Biofuels*, **11**, 201.
11. Ohm, R.A., Riley, R., Salamov, A., Min, B., Choi, I.-G. and Grigoriev, I.V. 2014, Genomics of wood-degrading fungi, *Fungal Genet. Biol.*, **72**, 82–90.
12. Suzuki, H., MacDonald, J., Syed, K., et al. 2012, Comparative genomics of the white-rot fungi, *Phanerochaete carnosa* and *P. chrysosporium*, to elucidate the genetic basis of the distinct wood types they colonize, *BMC Genomics*, **13**, 444.
13. Dai, Y., Li, X., Song, B., et al. 2019, Genomic analyses provide insights into the evolutionary history and genetic diversity of Auricularia species, *Front. Microbiol.*, **10**, 2255.
14. Lomascolo, A., Cayol, J.-L., Roche, M., et al. 2002, Molecular clustering of *Pycnoporus* strains from various geographic origins and isolation of monokaryotic strains for laccase hyperproduction, *Mycol. Res.*, **106**, 1193–203.
15. Nobles, M.K. and Frew, B.P. 1962, Studies in wood-inhabiting *Hymenomycetes* V. The genus *Pycnoporus* Karst, *Can. J. Bot.*, **40**, 987–1016.
16. Ryvarden, L. and Johansen, I. 1980, *A Preliminary Polypore Flora of East Africa*. Oslo: Fungiflora.
17. Justo, A. and Hibbett, D.S. 2011, Phylogenetic classification of Trametes (*Basidiomycota*, *Polyporales*) based on a five-marker dataset, *Taxon*, **60**, 1567–83.
18. Gnerre, S., MacCallum, I., Przybylski, D., et al. 2011, High-quality draft assemblies of mammalian genomes from massively parallel sequence data, *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 1513–8.
19. Miller, J.R., Delcher, A.L., Koren, S., et al. 2008, Aggressive assembly of pyrosequencing reads with mates, *Bioinformatics*, **24**, 2818–24.
20. Ka-Kit, L., LaButti, K., Khalak, A. and Tse, D. 2015, FinisherSC: a repeat-aware tool for upgrading de novo assembly using long reads, *Bioinformatics*, **31**, 3207–9.
21. Grigoriev, I.V., Nikitin, R., Haridas, S., et al. 2014, MycoCosm portal: gearing up for 1000 fungal genomes, *Nucleic Acids Res.*, **42**, D699–704.
22. Wattam, A.R., Abraham, D., Dalay, O., et al. 2014, PATRIC, the bacterial bioinformatics database and analysis resource, *Nucleic Acids Res.*, **42**, D581–91.

23. Katoh, K. and Standley, D.M. 2013, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, *Mol. Biol. Evol.*, **30**, 772–80.
24. Castresana, J. 2000, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, *Mol. Biol. Evol.*, **17**, 540–52.
25. Stamatakis, A. 2014, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, *Bioinformatics*, **30**, 1312–3.
26. Sima, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. 2015, Genome analysis BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, *Bioinformatics*, **31**, 3210–2.
27. Morin, E., Miyauchi, S., San Clemente, H., et al. 2019, Comparative genomics of *Rhizophagus irregularis*, *R. cerebriforme*, *R. diaphanus* and *Gigaspora rosea* highlights specific genetic features in Glomeromycotina, *New Phytol.*, **222**, 1584–98.
28. Miyauchi, S. 2020, Visual data pipeline: Proteomic Information Navigated Genomic Outlook (PRINGO). https://github.com/ShingoMiyauchi/PRINGO.
29. Auguie, B. 2017, Overview of the egg package. https://cran.r-project.org/web/packages/egg/index.html.
30. Yu, G., Smith, D.K., Zhu, H., Guan, Y. and Lam, T.T.-Y. 2017, ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data, *Methods Ecol. Evol.*, **8**, 28–36.
31. Wickham, H. 2009, *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
32. Kues, U., Nelson, D.R., Liu, C., et al. 2015, Genome analysis of medicinal *Ganoderma* spp. with plant-pathogenic and saprotrophic life-styles, *Phytochemistry*, **114**, 18–37.
33. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. and Henrissat, B. 2014, The carbohydrate-active enzymes database (CAZy) in 2013, *Nucleic Acids Res.*, **42**, D490–5.
34. Mitchell, A., Chang, H.-Y., Daugherty, L., et al. 2015, The InterPro protein families database: the classification resource after 15 years, *Nucleic Acids Res.*, **43**, D213–21.
35. Rawlings, N.D., Barrett, A.J. and Finn, R. 2016, Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors, *Nucleic Acids Res.*, **44**, D343–50.
36. Meux, E., Morel, M., Lamant, T., et al. 2013, New substrates and activity of *Phanerochaete chrysosporium* Omega glutathione transferases, *Biochimie*, **95**, 336–46.
37. Mathieu, Y., Prosper, P., Favier, F., et al. 2013, Diversification of fungal specific Class A glutathione transferases in saprotrophic fungi, *PLoS One*, **8**, e80298.
38. Roret, T., Thuillier, A., Favier, F., Gelhaye, E., Didierjean, C. and Morel-Rouhier, M. 2015, Evolutionary divergence of Ure2pA glutathione transferases in wood degrading fungi, *Fungal Genet. Biol.*, **83**, 103–12.
39. Pellegrin, C., Morin, E., Martin, F.M. and Veneault-Fourrey, C. 2015, Comparative analysis of secretomes from ectomycorrhizal fungi with an emphasis on small-secreted proteins, *Front. Microbiol.*, **6**, 1278.
40. Herpoël, I., Moukha, S., Lesage-Meessen, L., Sigoillot, J.-C. and Asther, M. 2000, Selection of *Pycnoporus cinnabarinus* strains for laccase production, *FEMS Microbiol. Lett.*, **183**, 301–6.
41. Navarro, D., Rosso, M.-N., Haon, M., et al. 2014, Fast solubilization of recalcitrant cellulosic biomass by the basidiomycete fungus *Laetisaria arvalis* involves successive secretion of oxidative and hydrolytic enzymes, *Biotechnol. Biofuels*, **7**, 143.
42. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. 2013, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, *Genome Biol.*, **14**, R36.
43. Love, M.I., Huber, W. and Anders, S. 2014, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biol.*, **15**, 550.
44. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. and Storey, J.D. 2012, The SVA package for removing batch effects and other unwanted variation in high-throughput experiments, *Bioinformatics*, **28**, 882–3.
45. Bolstad, B. 2019, preprocessCore: a collection of pre-processing functions. R package version 1.46.0. https://github.com/bmbolstad/preprocessCore.
46. Wehrens, R. and Buydens, L.M.C. 2007, Self- and super-organizing maps in R: the Kohonen Package, *J. Stat. Softw.*, **21**, 1–19.
47. Miyauchi, S., Navarro, D., Grigoriev, I.V., et al. 2016, Visual comparative omics of fungi for plant biomass deconstruction, *Front. Microbiol.*, **7**, 1335.
48. Miyauchi, S., Navarro, D., Grisel, S., Chevret, D., Berrin, J.-G. and Rosso, M.-N. 2017, The integrative omics of white-rot fungus *Pycnoporus coccineus* reveals co-regulated CAZymes for orchestrated lignocellulose breakdown, *PLoS One*, **12**, e0175528–e0175528.
49. Emms, D.M. and Kelly, S. 2019, OrthoFinder: phylogenetic orthology inference for comparative genomics, *Genome Biol.*, **20**, 238.
50. Lesage-Meessen, L., Haon, M., Uzan, E., et al. 2011, Phylogeographic relationships in the polypore fungus *Pycnoporus* inferred from molecular data, *FEMS Microbiol. Lett.*, **325**, 37–48.
51. Hori, C., Gaskell, J., Igarashi, K., et al. 2014, Temporal alterations in the secretome of the selective ligninolytic fungus *Ceriporiopsis subvermispora* during growth on aspen wood reveal this organism's strategy for degrading lignocellulose, *Appl. Environ. Microbiol.*, **80**, 2062–70.
52. Brown, N.A., Ries, L.N.A. and Goldman, G.H. 2014, How nutritional status signalling coordinates metabolism and lignocellulolytic enzyme secretion, *Fungal Genet. Biol.*, **72**, 48–63.
53. Alfaro, M., Oguiza, J.A., Ramírez, L. and Pisabarro, A.G. 2014, Comparative analysis of secretomes in basidiomycete fungi, *J. Proteomics*, **102**, 28–43.
54. Tandrup, T., Frandsen, K.E.H., Johansen, K.S., Berrin, J.-G. and Lo Leggio, L. 2018, Recent insights into lytic polysaccharide monooxygenases (LPMOs), *Biochem. Soc. Trans.*, **46**, 1431–47.
55. Henriksson, G., Johansson, G. and Pettersson, G. 2000, A critical review of cellobiose dehydrogenases, *J. Biotechnol.*, **78**, 93–113.
56. Berrin, J.-G., Rosso, M.-N. and Abou Hachem, M. 2017, Fungal secretomics to probe the biological functions of lytic polysaccharide monooxygenases, *Carbohydr. Res.*, **448**, 155–60.
57. Mäkelä, M.R., Marinović, M., Nousiainen, P., et al. 2014, Aromatic metabolism of filamentous fungi in relation to the presence of aromatic compounds in plant biomass, *Adv. Appl. Microbiol.*, **91**, 63–137.
58. Fernández-González, A.J., Valette, N., Kohler, A., et al. 2018, Oak extractive-induced stress reveals the involvement of new enzymes in the early detoxification response of *Phanerochaete chrysosporium*, *Environ. Microbiol.*, **20**, 3890–901.
59. Morel, M., Meux, E., Mathieu, Y., et al. 2013, Xenomic networks variability and adaptation traits in wood decaying fungi, *Microb. Biotechnol.*, **6**, 248–63.
60. Thuillier, A., Chibani, K., Belli, G., et al. 2014, Transcriptomic responses of *Phanerochaete chrysosporium* to oak acetonic extracts: focus on a new glutathione transferase, *Appl. Environ. Microbiol.*, **80**, 6316–27.
61. Mathieu, Y., Prosper, P., Buée, M., et al. 2012, Characterization of a *Phanerochaete chrysosporium* glutathione transferase reveals a novel structural and functional class with ligandin properties, *J. Biol. Chem.*, **287**, 39001–11.
62. Doddapaneni, H. and Yadav, J.S. 2005, Microarray-based global differential expression profiling of P450 monooxygenases and regulatory proteins for signal transduction pathways in the white rot fungus *Phanerochaete chrysosporium*, *Mol. Genet. Genomics*, **274**, 454–66.
63. Syed, K., Shale, K., Pagadala, N.S. and Tuszynski, J. 2014, Systematic identification and evolutionary analysis of catalytically versatile cytochrome p450 monooxygenase families enriched in model basidiomycete fungi, *PLoS One*, **9**, e86683.
64. Moktali, V., Park, J., Fedorova-Abrams, N.D., et al. 2012, Systematic and searchable classification of cytochrome P450 proteins encoded by fungal and oomycete genomes, *BMC Genomics*, **13**, 525.
65. Qhanya, L.B., Matowane, G., Chen, W., et al. 2015, Genome-wide annotation and comparative analysis of cytochrome P450 monooxygenases in *Basidiomycete* biotrophic plant pathogens, *PLoS One*, **10**, e0142100.
66. Ichinose, H. 2013, Cytochrome P450 of wood-rotting basidiomycetes and biotechnological applications, *Biotechnol. Appl. Biochem.*, **60**, 71–81.

67. Song, J. and Cui, B.-K. 2017, Phylogeny, divergence time and historical biogeography of *Laetiporus* (*Basidiomycota*, *Polyporales*), *BMC Evol. Biol.*, **17**, 102.
68. Castanera, R., López-Varas, L., Borgognone, A., et al. 2016, Transposable elements versus the fungal genome: impact on whole-genome architecture and transcriptional profiles, *PLoS Genet.*, **12**, e1006108.
69. Nagy, L.G., Riley, R., Bergmann, P.J., et al. 2017, Genetic bases of fungal white rot wood decay predicted by phylogenomic analysis of correlated gene-phenotype evolution, *Mol. Biol. Evol.*, **34**, 35–44.
70. Garajova, S., Mathieu, Y., Beccia, M.R., et al. 2016, Single-domain flavoenzymes trigger lytic polysaccharide monooxygenases for oxidative degradation of cellulose, *Sci. Rep.*, **6**, 28276.
71. Kracher, D., Scheiblbrandner, S., Felice, A.K.G., et al. 2016, Extracellular electron transfer systems fuel cellulose oxidative degradation, *Science*, **352**, 1098–101.
72. Bissaro, B., Várnai, A., Røhr, Å.K. and Eijsink, V.G.H. 2018, Oxidoreductases and reactive oxygen species in conversion of lignocellulosic biomass, *Microbiol. Mol. Biol. Rev.*, **82**, e00029.
73. Sützl, L., Laurent, C.V.F.P., Abrera, A.T., Schütz, G., Ludwig, R. and Haltrich, D. 2018, Multiplicity of enzymatic functions in the CAZy AA3 family, *Appl. Microbiol. Biotechnol.*, **102**, 2477–92.
74. Hyde, S.M. and Wood, P.M. 1997, A Mechanism for production of hydroxyl radicals by the brown-rot fungus *Coniophora puteana*: Fe(III) reduction by cellobiose dehydrogenase and Fe(II) oxidation at a distance from the hyphae, *Microbiology*, **143**, 259–66.
75. Tan, T.-C., Kracher, D., Gandini, R., et al. 2015, Structural basis for cellobiose dehydrogenase action during oxidative cellulose degradation, *Nat. Commun.*, **6**, 7542.
76. Bissaro, B., Røhr, Å.K., Müller, G., et al. 2017, Oxidative cleavage of polysaccharides by monocopper enzymes depends on H_2O_2, *Nat. Chem. Biol.*, **13**, 1123–8.
77. Kersten, P. and Cullen, D. 2014, Copper radical oxidases and related extracellular oxidoreductases of wood-decay Agaricomycetes, *Fungal Genet. Biol.*, **72**, 124–30.
78. Fernández-Fueyo, E., Ruiz-Dueñas, F.J., Miki, Y., Martínez, M.J., Hammel, K.E. and Martínez, A.T. 2012, Lignin-degrading peroxidases from genome of selective ligninolytic fungus *Ceriporiopsis subvermispora*, *J. Biol. Chem.*, **287**, 16903–16.
79. Salvachúa, D., Martínez, A.T., Tien, M., et al. 2013, Differential proteomic analysis of the secretome of *Irpex lacteus* and other white-rot fungi during wheat straw pretreatment, *Biotechnol. Biofuels*, **6**, 115.
80. Fernández-Fueyo, E., Ruiz-Dueñas, F.J., López-Lucendo, M.F., et al. 2016, A secretomic view of woody and nonwoody lignocellulose degradation by *Pleurotus ostreatus*, *Biotechnol. Biofuels*, **9**, 49.
81. Kuuskeri, J., Häkkinen, M., Laine, P., et al. 2016, Time-scale dynamics of proteome and transcriptome of the white-rot fungus *Phlebia radiata*: growth on spruce wood and decay effect on lignocellulose, *Biotechnol. Biofuels*, **9**, 192.
82. Moody, S.C., Dudley, E., Hiscox, J., Boddy, L. and Eastwood, D.C. 2018, Interdependence of primary metabolism and xenobiotic mitigation characterizes the proteome of *Bjerkandera adusta* during wood decomposition, *Appl. Environ. Microbiol.*, **84**, e01401–17.
83. Kersten, P.J. 1990, Glyoxal oxidase of *Phanerochaete chrysosporium*: its characterization and activation by lignin peroxidase, *Proc. Natl. Acad. Sci. U. S. A.*, **87**, 2936–40.
84. Martinez, D., Challacombe, J., Morgenstern, I., et al. 2009, Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion, *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 1954–9.
85. Daou, M. and Faulds, C.B. 2017, Glyoxal oxidases: their nature and properties, *World J. Microbiol. Biotechnol.*, **33**, 87.

IV.2. Gene family expansions and transcriptome signatures uncover fungal adaptations to wood decay.

Chapter IV.2 summary

Due to their unique ability to degrade all constituents of wood, including lignin, wood decay fungi are the ultimate decomposers of dead organic matter in forest ecosystems. As such, they play a role in the dynamics of carbon cycling. In addition, on a biotechnological viewpoint, they do contribute to innovations in the bioeconomy, as a source of enzymatic biocatalysts for plant biomass conversion into bio-based fibers, chemicals and biofuels.

Fungi from the order Polyporales are among the most efficient decomposers of dead organic matter in forest ecosystems. Previous studies suggest that Polyporales species use diverse enzymatic mechanisms during wood decay. Yet, the vision on their evolutionary adaptation to wood decay and genome diversity remains incomplete.

After exploring the intra-genus diversity in the previous chapter, our objective in this study was to assess the diversity related to lignocellulose machineries at a higher taxonomic level, i.e. between species from the Polyporales order, and to identify the genomic features that supported adaption to wood decay in these fungi.

For this purpose, the genomes of 26 Polyporales species have been newly sequenced, covering the main phylogenetic clades of Polyporales. We analyzed these genomes together with 24 previously published genomes. Part of this analysis was done during my visit to the INRAE lab in Nancy “Interactions Arbres-Microorganisms”.

We estimated that Polyporales might have diverged 183 million years ago and highlighted that during this long evolutionary period they have undergone intensive diversification. Indeed, we observed poor synteny conservation, suggesting that genome rearrangements have shaped Polyporales genomes. Besides, using phylogenomics, we spotted different gene family expansion/contraction trajectories for genes coding for plant cell wall degrading enzymes and for genes involved in signaling and regulation among the different clades. Those diverse evolutionary trends have shaped the gene portfolios in extant species, despite Polyporales mainly comprise wood decayers.

In addition, using cross-species comparative transcriptomics, we identified yet overlooked conserved genes that code for enzymes active at the fungal cell wall-plant cell wall interface and could be involved in the degradation of recalcitrant plant cell wall polymers.

These findings highlight the functional diversity observed within a single order, the Polyporales, which may have contributed to their adaptation to different wood substrates and ecological niches and which therefore could help in improving the design of biocatalysts aimed at the use of recalcitrant plant materials as a renewable carbon source.

This work is presented in the manuscript “Gene family expansions and transcriptome signatures uncover fungal adaptations to wood decay”, by Hayat Hage, Shingo Miyauchi, Máté Virágh, Elodie Drula, Byoungnam Min, Delphine Chaduli, David Navarro, Anne Favel, Manon Norest, Laurence Lesage-Meessen, Balázs Bálint, Zsolt Merényi, Laura de Eugenio, Emmanuelle

Morin, Angel T. Martínez, Petr Baldrian, Martina Štursová, María Jesús Martínez, Cenek Novotny, Jon K. Magnuson, Joey W. Spatafora, Sundy Maurice, Jasmyn Pangilinan, Willian Andreopoulos, Kurt LaButti, Hope Hundley, Hyunsoo Na, Alan Kuo, Kerrie Barry, Anna Lipzen, Bernard Henrissat, Robert Riley, Steven Ahrendt, László G. Nagy, Igor V. Grigoriev, Francis Martin and Marie-Noëlle Rosso, submitted for publication.

Supplementary tables are available at the following link:
https://figshare.com/s/144d7ab3ae2ef31b1db3

Gene family expansions and transcriptome signatures uncover fungal adaptations to wood decay.

Running title: Evolutionary routes for wood decay enzyme toolkits

Hayat Hage[1]*, Shingo Miyauchi[1,$]*, Máté Virágh[2], Elodie Drula[1,3], Byoungnam Min[4,5], Delphine Chaduli[1,6], David Navarro[1,6], Anne Favel[1,6], Manon Norest[1], Laurence Lesage-Meessen[1,6], Balázs Bálint[2], Zsolt Merényi[2], Laura de Eugenio[7], Emmanuelle Morin[8], Angel T. Martínez[7], Petr Baldrian[9], Martina Šturcová[9], María Jesús Martínez[7], Cenek Novotny[9,10], Jon K. Magnuson[11], Joey W. Spatafora[12], Sundy Maurice[13], Jasmyn Pangilinan[4], Willian Andreopoulos[4], Kurt LaButti[4], Hope Hundley[4], Hyunsoo Na[4], Alan Kuo[4], Kerrie Barry[4], Anna Lipzen[4], Bernard Henrissat[14], Robert Riley[4], Steven Ahrendt[4], László G. Nagy[2,15], Igor V. Grigoriev[4,5,16], Francis Martin[8], Marie-Noëlle Rosso[1,§].

1 INRAE, Aix Marseille Univ, UMR1163, Biodiversité et Biotechnologie Fongiques, 13009, Marseille, France

2 Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Center, 6726, Szeged, Hungary

3 INRAE, USC1408, AFMB, 13009, Marseille, France

4 US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

5 Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

6 INRAE, Aix Marseille Univ, CIRM-CF, UMR1163, 13009, Marseille, France

7 Centro de Investigaciones Biológicas Margarita Salas, CIB-CSIC, 28040, Madrid, Spain

8 Université de Lorraine, INRAE, UMR1136, Interactions Arbres/Microorganismes, 54280, Champenoux, France

9 Institute of Microbiology of the Czech Academy of Sciences, 142 20 Praha 4, Czech Republic

10 University of Ostrava, 701 03 Ostrava, Czech Republic

11 Pacific Northwest National Laboratory, Richland, WA 99352, USA

12 Department of Botany and Plant Pathology, Oregon State University, Corvallis, 97331 OR, USA

13 Section for Genetics and Evolutionary Biology, University of Oslo, 0316, Oslo, Norway

14 Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia

15 Department of Plant Anatomy, Institute of Biology, Eötvös Loránd University, Budapest, 1117, Hungary

16 Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA, USA

$ present address: Max Planck Institute for Plant Breeding Research, Department of Plant Microbe Interactions, Köln, Germany

* These two authors contributed equally as first author

§ corresponding author: Dr. Marie-Noëlle Rosso, Bât Polytech entrée B, 163 Av de Luminy, Case 925, 13288 Marseille Cedex 9 ; Tél : +(33)0 4 91 82 86 07 ; Fax : +(33)0 4 91 82 86 01, e-mail : marie-noelle.rosso@inrae.fr

Abstract

Because they comprise some of the most efficient wood-decayers, Polyporales fungi impact carbon cycling in forest environment. Despite continuous discoveries on the enzymatic machinery involved in wood decomposition, the vision on their evolutionary adaptation to wood decay and genome diversity remains incomplete.

We combined the genome sequence information from 50 Polyporales species, including 26 newly sequenced genomes and sought for genomic and functional adaptations to wood decay through the analysis of genome composition and transcriptome responses to different carbon sources.

The genomes of Polyporales from different phylogenetic clades showed poor conservation in macrosynteny, indicative of genome rearrangements. We observed different gene family expansion/contraction histories for plant cell wall degrading enzymes in core polyporoids and phlebioids and captured expansions for genes involved in signaling and regulation in the lineages of white rotters. Furthermore, we identified conserved cupredoxins, thaumatin-like proteins and Lytic Polysaccharide Monooxygenases with a yet uncharacterized appended module as new candidate players in wood decomposition.

Given the current need for enzymatic toolkits dedicated to the transformation of renewable carbon sources, the observed genomic diversity among Polyporales strengthens the relevance of mining Polyporales biodiversity to understand the molecular mechanisms of wood decay.

Keywords

Biodiversity, CAZyme, phylogenomics, fungal ecology, lignocellulose, plant cell wall, wood decay.

Introduction

The Polyporales order, which represents 1.5% of all described fungal species (Kirk *et al.*, 2009), belongs to the Agaricomycetes class, a highly diverse and large group of mushroom-forming fungi within the phylum Basidiomycota. Polyporales are widely distributed in boreal, temperate, tropical and subtropical areas. The great majority of Polyporales are efficient wood decay saprotrophs with significant contribution to carbon re-allocation and carbon cycling in forest soils (Boberg *et al.*, 2014; Chen *et al.*, 2019). Based on decay symptoms the fungal attack causes to the wood, Polyporales and other wood decay fungi have been classified traditionally into two rotting types: white rotters, which degrade all plant cell wall (PCW) polymers, including lignin, and leave lighter-colored cellulose behind, and brown rotters, which degrade the PCW cellulosic and hemicellulosic compounds with partial modification of lignin, resulting in brown cubic wood fragments (Blanchette, 1984; Martínez *et al.*, 2005). However, beyond these archetypal situations, a continuum in the ratio of lignin vs. carbohydrate depolymerization during wood colonization has been observed, which highlights diverse strategies for lignocellulose modifications (Schilling *et al.*, 2020). Besides saprotrophs, additional nutrition modes have been observed for some species. The pathogens colonize living trees through wounds and cause wood rot disease. More occasionally, endophytes in leaves and sapwood of trees, are suggested to stand as latent saprotrophs (Sokolski *et al.*, 2007; Martin *et al.*, 2015). Finally, some have been described as potential temporary parasites of other fungi preliminarily established in the wood (Rayner *et al.*, 1987).

The order Polyporales has been subdivided phylogenetically into two lineages including, among others, the core polyporoid, gelatoporia, and antrodia clades on the one hand, and the phlebioid and residual polyporoid clades on the other hand (Justo *et al.*, 2017). Most species produce white-rot, the plesiomorphic decay type for Polyporales (Floudas *et al.*, 2012), except species from the antrodia lineage and a few genera that convergently transitioned to the brown-rot type (Justo *et al.*, 2017).

Each rot type involves a characteristic panel of enzymatic and chemical mechanisms for the degradation of PCW polymers (reviewed in (Lundell *et al.*, 2014)), which is generally reflected by a distinct repertoire of genes coding PCW degrading enzymes (PCWDE) in the genome. Notably, carbohydrate-active enzymes (CAZymes) related to the depolymerization of crystalline cellulose such as Lytic Polysaccharide Monooxygenases (LPMO) or to the depolymerization of lignin (Class II peroxidases, POD) and auxiliary enzymes such as multicopper oxidases (MCO) and H_2O_2-generating glucose–methanol–choline-oxidoreductases are generally expanded in the genome of white-rot fungi (Floudas *et al.*, 2012; Ruiz-Dueñas *et al.*, 2013; Nagy *et al.*, 2016; Miyauchi *et al.*, 2020), although in some species, the absence of genes for lignin-active POD does not fit the white-rot phenotype (Riley *et al.*, 2014). Conversely, brown-rot fungi largely rely on the production of reactive oxygen species (ROS) via the iron-dependent Fenton chemistry for the

oxidative attack of wood polymers prior to enzymatic degradation (Martínez *et al.*, 2005; Eastwood *et al.*, 2011). Accordingly, the transition from white-rot to brown-rot was associated with the loss of key POD enzymes for lignin degradation (Floudas *et al.*, 2012) and a change in the temporal expression of oxidative vs. hydrolytic enzymes in the early stages of wood degradation (Zhang *et al.*, 2019). Besides PCW degradation, other physiological functions are involved in the ability to grow on wood, such as the intracellular transport and detoxification of toxic products released during the degradation process, as reflected by enrichment in the gene portfolios for Major Facilitator Superfamily transporters, cytochrome P450 and glutathione transferases in the genomes of these fungi (Syed *et al.*, 2014; Deroy *et al.*, 2015; Qhanya *et al.*, 2015; Nagy *et al.*, 2016).

Beyond those global characteristics, white-rot and brown-rot fungi each developed a diversity of enzymatic mechanisms to cope with different PCW chemical compositions (Suzuki *et al.*, 2012; Deroy *et al.*, 2015; Presley & Schilling, 2017). Within Polyporales, white-rot species degrade lignin with different degrees of cellulose preservation (Fernandez-Fueyo *et al.*, 2012; Miyauchi *et al.*, 2018) and use different sets of enzymes to degrade lignin. For example, within the phlebioid clade, some species have no laccase, an enzyme with ligninolytic activity (Rivera-Hoyos *et al.*, 2013), (e.g *Phanerochaete chrysosporium*, *Phanerochaete carnosa*; Suzuki et al., 2012) whereas four and five laccase genes have been identified in *Phlebia centrifuga* and *Phlebia radiata* respectively (Kuuskeri *et al.*, 2016). Similarly, in the genomes of core polyporoids, the number of lignin-active PODs varies significantly from one species to another. For example, no genes of lignin peroxidase, a subfamily of PODs, were identified in the genome of *Dichomitus squalens* whereas ten genes were found in that of *Trametes versicolor* (Ruiz-Dueñas *et al.*, 2013).

This diversity in enzymatic systems in Polyporales is of high interest to address the current demand for bio-based chemicals and materials derived from renewable carbon source – the plant biomass – without competition with food or feed production or with the preservation of natural habitats. In this context, the plant tissues to be transformed are issued from wastes or co-products of agriculture and forestry, which are recalcitrant to degradation. The recent progress in understanding the enzymatic mechanisms used by Polyporales for the oxidative degradation of non-starch polymers (Zhou *et al.*, 2015; Berrin *et al.*, 2017; Martínez *et al.*, 2017; Couturier *et al.*, 2018) has stimulated the screening of Polyporales strains for process development and the identification of strains able to use raw biomasses as carbon source (Berrin *et al.*, 2012; Zhou *et al.*, 2015; Miyauchi *et al.*, 2018).

A recent phylogenomic study suggested that high diversification rates occurred in Polyporales (Varga *et al.*, 2019), which could have shaped enzymatic mechanisms to adapt diverse ecological niches and substrates, and further strengthens the relevance of the exploration of this taxon for enzymatic diversity. Yet, further studies are required to extend our

understanding on the genomic adaptations to wood decay and to assess the functional diversity of the enzymatic systems across Polyporales.

In this work, we investigated the genomic basis of evolutionary adaptations related to wood decay within Polyporales. We compared the phylogenomic features and the repertoires of protein-coding genes of 50 fungal species including 26 newly sequenced genomes. The genomic diversity within the order was examined in terms of their genome arrangement and the evolution of gene families. The transcriptomic trends of selected Polyporales species from the core polyporoid and phlebioid clades during degradation of diverse lignocellulosic substrates led to the discovery of conserved gene sets regulated for PCW degradation. Our results unveil some of the mechanisms underlying Polyporales diversification and pinpoint to yet overlooked proteins that could contribute to the ability of Polyporales to degrade recalcitrant PCW polymers.

Results

Polyporales phylogeny and genome features

We sequenced the genome of 26 species, including two species (*Abortiporus biennis* and *Panus rudis*) from the residual polyporoid clade, for which no genome was available, and species from the core polyporoid (16 genomes), antrodia (4 genomes) and phlebioid clades (4 genomes; Table S1, Table S2). We used these 26 genomes along with 24 previously published Polyporales genomes in a phylogenetic reconstruction of Polyporales based on 540 single-copy genes (Table S3). The topology of the maximum likelihood tree confirmed the phylogeny based on three genes previously reported by Justo et al. (2017) and the classification into the five main clades; core polyporoid, antrodia (grouping all brown-rot Polyporales), gelatoporia, phlebioid and residual polyporoid. The tree topology comforted the placement of the residual polyporoid clade as the sister of the phlebioid clade (Fig. 1a). We estimated the divergence times of the major Polyporales lineages. Molecular clock analyses of a phylogenomic dataset showed that the Polyporales might be around 183 million years old, which is in line with recent molecular clock studies of Agaricomycetes (Varga *et al.*, 2019). Trametoid species were estimated at ca. 61 myr, which is consistent with a 45 myr old Trametes fossil from the literature (Knobloch & Kotlaba, 1994).

Overall, our study included 36 white rotters and ten brown rotters, including *Laetiporus sulphureus*, which has been described both as a brown rotter and a pathogen (Song & Cui, 2017), and four saprotroph species with no attributed decay type (Fig. 1b). The assembly size of the 50 Polyporales genomes ranged from 28 to 66 Mb (mean 39 Mb, standard deviation SD 7) and the transposable element (TE) coverage was positively correlated to genome size (0.37% to

16% TE coverage, mean 8%, SD 8, Pearson correlation coefficient = 0.4, p-value 0.001). The most abundant TEs were Long Terminal Repeats from the Gypsy and Copia families (Fig. S1).

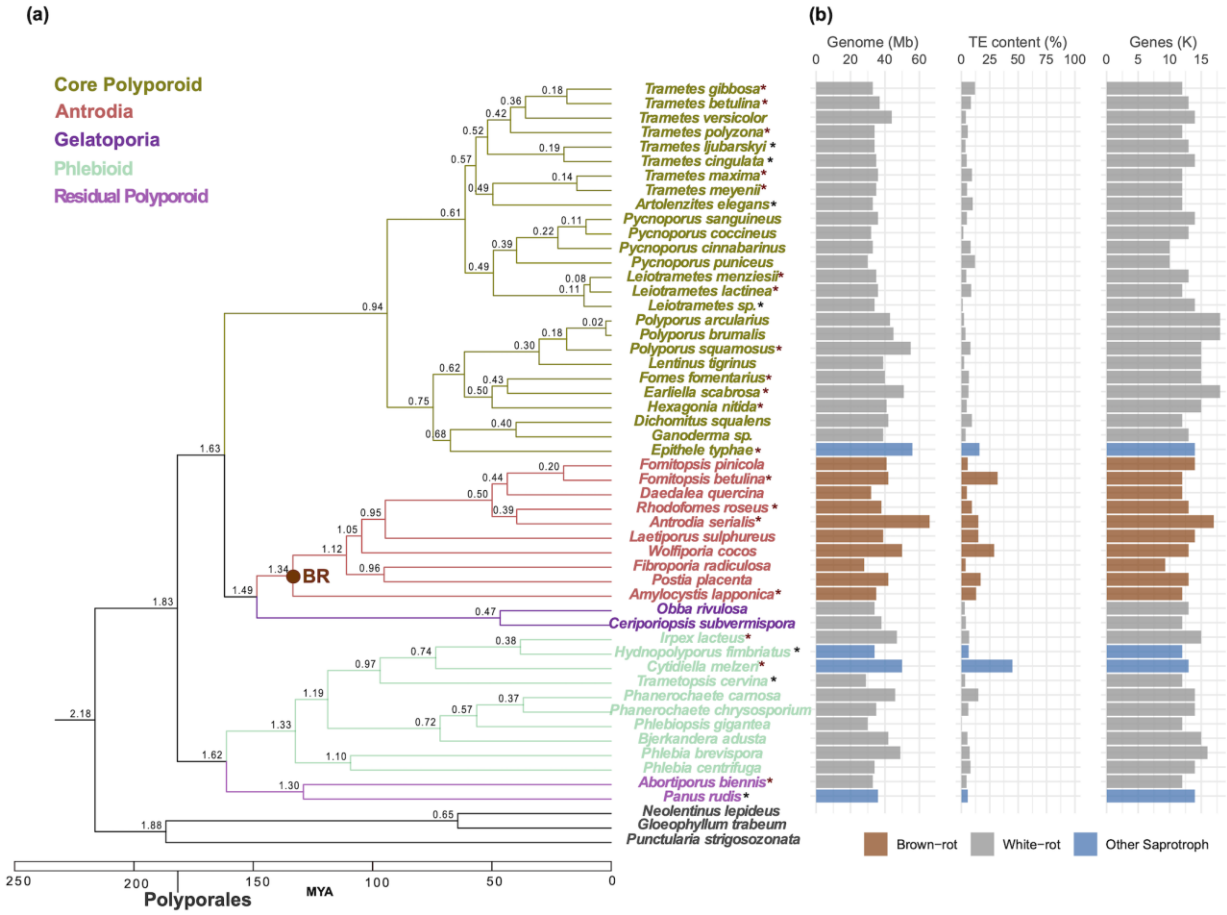


Figure 1 Phylogenetic relationship and genome features of Polyporales fungi. (a) Time-calibrated species phylogeny. Numbers next to nodes represent mean estimated ages from MCMCTree using six fossil calibrations. (b) Genome size, genome content in transposable elements (TE) and number of protein-coding genes. The newly sequenced genomes are indicated with asterisks. Genomes sequenced with long read methods are indicated with red asterisks.

To assess the conservation of the genome structure across Polyporales, we analyzed the genome synteny within and across clades. For this analysis we selected the 23 Polyporales genomes that had their 10 largest scaffolds covering more than 40% of the genome (Table S4). We quantified synteny by calculating the percentage of hits occurring in the same order on the compared sequences. The genomes from the core polyporoid clade showed a higher synteny conservation (mean 13.5% hits in synteny) than those from the antrodia (7.6%) or phlebioid clades (7.9%) and the synteny dropped significantly between genomes from different clades (Fig 2; Fig. S2). The higher macrosynteny in core polyporoid genomes might be related to this clade being younger (94 Mya; Fig. 1a) compared to the antrodia and phlebioid clades (estimated ~134 and ~133 Mya). It is noteworthy that no significant correlation was observed between TE

coverage and synteny loss (Pearson correlation coefficient = -0.12, p-value 0.2), suggesting that TE movements were not the major drivers for genome reorganizations.

Core, dispensable and species-specific genes

The protein-coding genes from the 50 Polyporales genomes were clustered in orthologous groups and used to calculate the counts of core, dispensable and species-specific genes. The core genes conserved among all genomes consisted of 2,169 clusters and its proportion in each species varied from 13 to 26% (mean 18%, SD 2) of the total proteome. The proportion of species-specific genes varied from 11 to 43% (mean 26%, SD 8). From the latter, 97% (mean 98%, SD 5) had no homologs in all annotated fungal genomes in MycoCosm (as of Oct 2019, 1420 genomes) (Fig. S3; Table S5). Intriguingly, we identified that an average of 81% of these species-unique genes were annotated as small secreted proteins (<300 aa, SSP; Fig. S4, S5). To test whether the predicted SSP genes could result from defects in gene sequencing and assembly or from defects in gene structural annotations, we looked at transcript evidence in the transcriptomes of six species, *Irpex lacteus*, *Artolenzites elegans*, *Leiotrametes* sp., *Pycnoporus cinnabarinus*, *Pycnoporus coccineus* and *Trametes ljubarskyi* (see below). In each species, in our growth conditions, we identified transcripts for 87 to 94% of the predicted SSP genes, comforting the reliability of our SSP gene predictions.

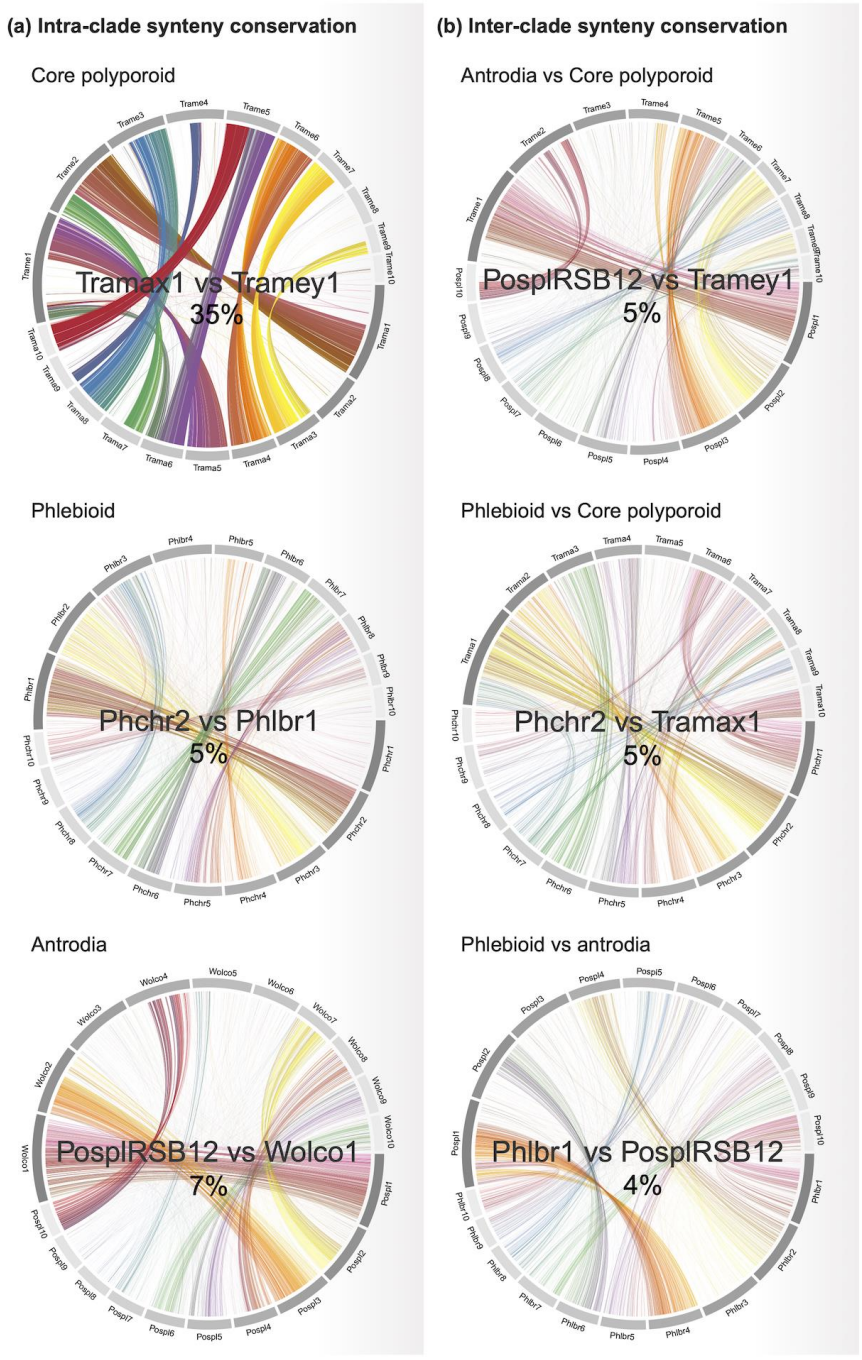


Figure 2 Comparison of synteny conservation between genomes from different Polyporales clades: core polyporoid (Trametes meyenii vs. Trametes maxima), phlebioid (Phanerochaete chrysosporium vs. Phlebia brevispora) and antrodia (Postia placenta vs. Wolfiporia cocos). The percentage of hits occurring in the same order on the compared block of sequences between two genomes is shown inside each circos plot.

The phylogenetic clades within Polyporales have different portfolios of secreted CAZymes

We assessed the diversity of the gene repertoires for enzymes involved in lignocellulose degradation across Polyporales, by comparing the repertoires of genes coding for secreted CAZymes in the 50 genomes. Phylogeny had a higher impact than rot type on the repertoires of secreted CAZymes in Polyporales [38% and 7% of the variance respectively, PERMANOVA (Mesny, 2020)]. As expected, we observed that the gene repertoires of brown rotters (from the antrodia clade) were distinct from those of white rotters (Fig. 3a, Fig. S6), mainly due to significantly reduced numbers of PCWDE in antrodia genomes, notably laccases (AA1_1), POD (AA2), carbohydrate binding module 1 (CBM1), AA9 LPMOs and expansins (Fig. 3b, Fig. S7; Table S6). Among white rotters, the phlebioids had reduced numbers of laccase genes (Kruskal-wallis test, p value < 0.05), whereas core polyporoids were enriched for GH18 chitinases and for GH79 β-glucuronosidases, which target 4-O-methyl-glucuronic acids from plant hemicelluloses (Kuroyama et al., 2001). Interestingly, the gelatoporia (white-rotters), a sister clade of antrodia (brown rotters), had similar gene counts as core polyporoids or phlebioids for PODs, CBM1 modules, AA5_1 and PL14_4, but resembled antrodia by low numbers of laccase, AA9 LPMO and expansin genes. Although the two gelatoporia species analyzed here (*Obba rivulosa* and *Ceriporiopsis subvermispora*) are white rot fungi (Blanchette *et al.*,1997; Miettinen *et al.*, 2016), this similitude with antrodia repertoires suggested gene losses had occurred in a common ancestor of the two clades (see below).

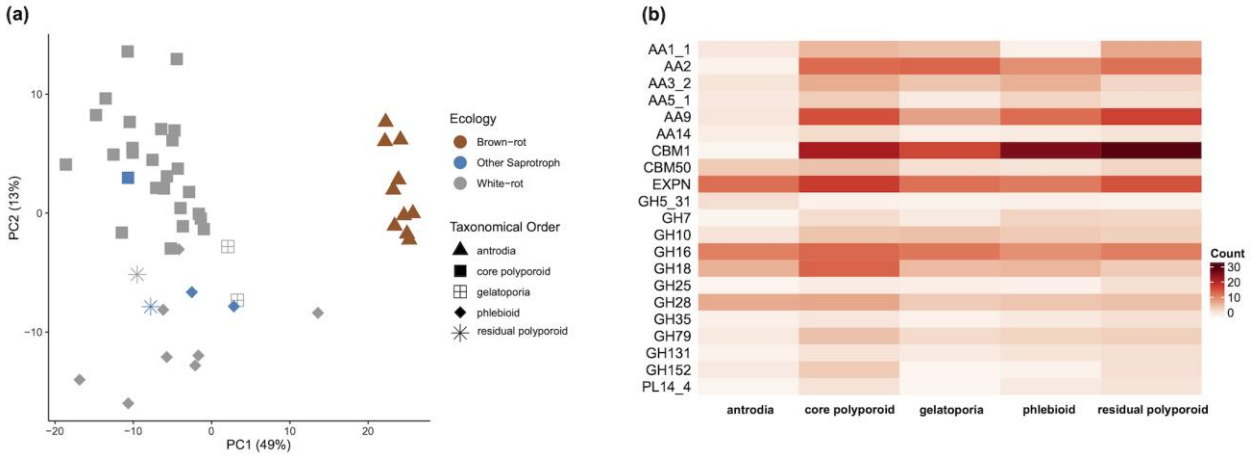


Figure 3 Comparison of the counts of secreted CAZyme genes in 50 Polyporales species. (a) PCA of the counts of genes coding for secreted CAZymes in 50 Polyporales species. (b) Heatmap showing the differences between Polyporales clades in counts of the secreted CAZymes contributing the most to the distribution of the species in the PCA plot.

CAZyme genes underwent different gain/loss trajectories in Polyporales lineages.

We further investigated the gene family expansion/contraction events that took place on eight early nodes of the Polyporales, partly because we wanted to avoid the effects of large, species-specific expansions that can stem from genome assembly artefacts and because early events are more likely to generate conserved genetic signatures of wood decay capabilities of the whole order. We found that extensive CAZyme gene loss took place before diversification of antrodia and gelatoporia, which was congruent with reduced numbers of AA9 and expansins in white rot gelatoporia (Fig. 4). Surprisingly, the MRCA of phlebioids had a smaller reconstructed ancestral CAZyme repertoire (531 genes) than that of the core polyporoid clade (645 genes). This is driven by a large difference in the number of gene duplications inferred in the MRCA of the core polyporoid vs that of the phlebioid clade: 103 duplications vs nine (in comparison, only six in the antrodia clade). We find that several oxidoreductases, CBM1-s and expansins contribute to this pattern. For example, the laccase, POD and GMC-oxidoreductase families showed four, three and 12 duplications in the MRCA of the core polyporoid clade, respectively, but 0, 0 and 5 duplications in the MRCA of the phlebioid clade, respectively. Similarly, CBM1-containing genes and expansins showed 17 and 13 duplications in the former while only three and two in the latter clade. LPMO-s (AA9) showed a more moderate difference (3 duplications in the MRCA of the core polyporoid vs one in that of the phlebioid). These differences in gene family expansions in the MRCA of Polyporoid clades have resulted in reduced numbers of laccase genes (AA1_1) in extant phlebioid genomes (Fig. S7). However, in other CAZyme families, our reconstruction suggests that successive gene loss events took place in core polyporoids whereas gene gains were predominant in phlebioids. For example, in core polyporoids, gene gains in general peroxidase (GP) were followed by extensive gene loss in several lineages (Fig. S8).

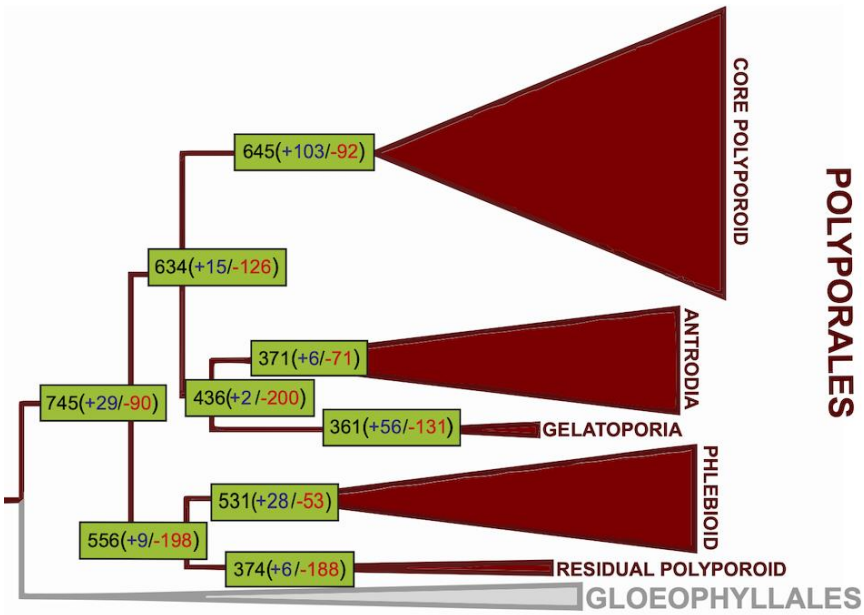


Figure 4 Numbers of genes coding for CAZymes in the most recent common ancestors of the five Polyporales clades.

We further inspected the diversity of POD gene portfolios among Polyporales. We classified the AA2 POD genes based on the presence of known conserved residues into the three POD types active on lignin; MnP, VP and LiP (Ruiz-Dueñas *et al.*, 2013). Our study on ten antrodia genomes confirmed previous observations on the presence of non-ligninolytic GP and the absence of MnP, LiP and VP in these brown rotters. Conversely, the white-rot species showed diverse combinations of the POD types. From the 40 genomes analyzed, 15 contained the three types of lignin-active peroxidases, 19 contained combinations of MnP and VP or MnP and LiP and six contained MnP only (Fig. 5). In addition, we noticed rare occurrences of atypical MnP with two acidic amino acids (QED and EQE) instead of three at the Mn(II) binding site, that were not previously described in Agaricomycetes (Ruiz-Dueñas *et al.*, 2020).

Gene tree-species tree reconciliation studies suggested that the diversity in POD portfolios among Polyporales resulted from different evolutionary histories in the different lineages, and from recent gene gain events for LiP genes in the phlebioids *Phanerochate chrysosporium* and *Bjerkandera adusta*, for MnP in gelatoporia (*Ceriporiopsis subvermispora*) and residual polyporoids (*Abortiporus biennis)*, and for VP in core polyporoids and in the residual polyporoid *Panus rudis* (Fig. S8).

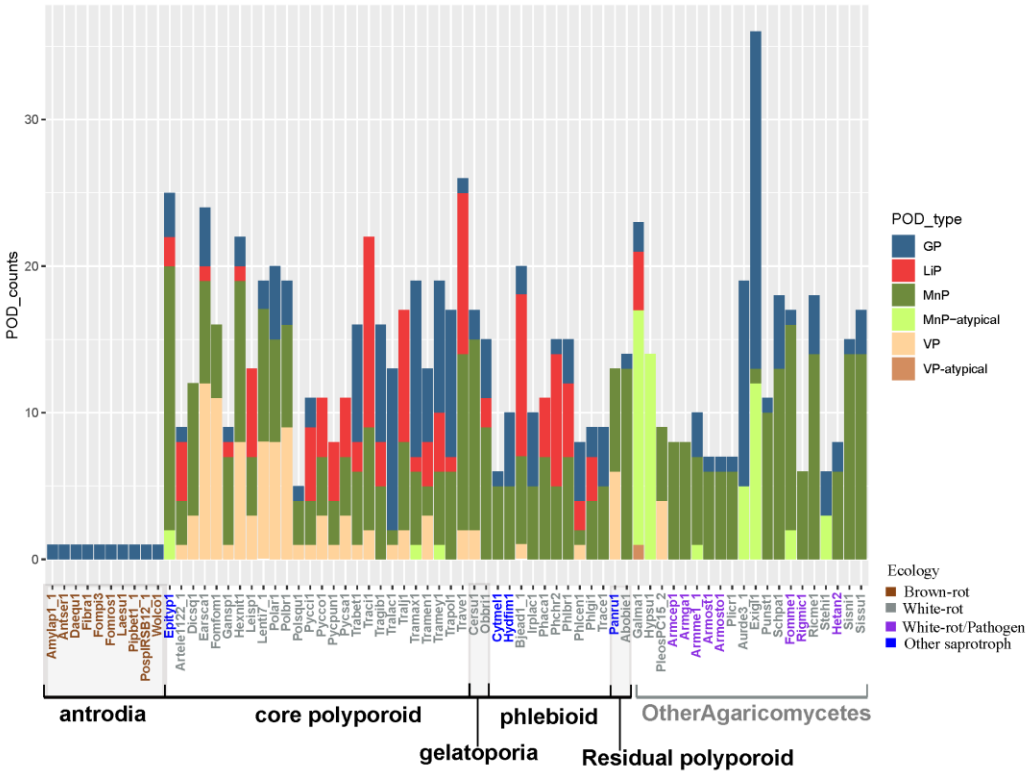


Figure 5 Count of Class II peroxidases (PODs) in Polyporales genomes and in Agaricomycetes species from other taxonomic orders.

Several gene families underwent intensive expansions in early Polyporales

We searched for gene families whose expansion/contraction trajectories had potential functional consequences upon major Polyporales clades, with no *a priori* knowledge on these functions. To answer this question, we searched which gene families expanded most significantly in basal nodes of the Polyporales and converted gene duplication statistics to duplication rates; this allowed us to rank gene families by their duplication rate in early Polyporales clades. The predicted functions for each cluster were deduced from InterPro domains (Table S7). The ten most significantly expanding clusters at the Polyporales MRCA had predicted functions related to signaling and protein-protein interactions. Cluster12444, which contains proteins with an EF-Hand, calcium binding domain (IPR018247), had the highest gain rate and showed expansions in seven of the eight examined early-Polyporales nodes suggesting that it is involved in key wood-decay processes. Both a survey of published RNASeq data (Miyauchi *et al.*, 2018, 2020; Wu *et al.*, 2018) and the data generated in this study showed that genes from this cluster were differentially expressed in seven out of eight inspected Polyporales species in response to diverse ligno(cellulosic) substrates (Table S8). Similarly, we found evidence for differential expression during growth on lignocellulosic substrates, for eight out of

the ten most significantly expanding clusters, which contained protein kinase domains (Table S8).

Expansion and contraction signatures of white-rot and brown-rot Polyporales

To refine our search for potentially decay-associated gene families, we constructed a model with the prerequisite that a gene family showed the following attributes: (i) expansion in early white-rot-associated nodes of the Polyporales, (ii) losses in the brown-rot antrodia clade and (iii) the family is conserved in >50% of white-rot species. We detected 338 gene families that fit this model (Table S9). Somewhat surprisingly, the most expanding families in the white-rot Polyporales included only a few PCWDEs, AA9 LPMOs (three clusters), a GH5 and a GH7 cellulase, a GH30 endo-1,4-β-xylanase, a GMC-oxidoreductase cluster, as well as PODs. The lack of expansion for other PCWDE is consistent with the inference that white-rot originated long before the Polyporales (Floudas *et al.*, 2012), so probably relevant PCWDE expansions also happened earlier. On the other hand, a preponderance of peptidase (13 clusters), cytochrome p450s (13 clusters) and transporter (13 clusters) clusters was obvious among the families which are most highly duplicated in early Polyporales but are lost in brown rotters (Table S9). Furthermore, several groups of regulator families, including seven clusters of F-box domain proteins, eight of transcription factors, two clusters of BTB/POZ proteins, four clusters of RING/MYND type zinc finger proteins (both of which can indirectly regulate transcription), and six of protein kinases were found. Based on their expansion/contraction patterns, we speculate that these families may be functionally linked to wood decay; they comprise worthy targets for detailed functional follow-up experiments.

Transcriptomics show conserved signatures for white-rot decay across Polyporales

To further assess the functional adaptation of Polyporales for wood decay activities, we compared the transcriptomic responses to three lignocellulosic substrates representing gramineae (wheat-straw), softwood (pine) or hardwood (aspen) using the following species, *Artolenzites elegans*, *Leiotrametes* sp., *Pycnoporus cinnabarinus*, *Pycnoporus coccineus*, *Trametes ljubarskyi* and *Irpex lacteus*. We used the differential expression level of the genes coding predicted secreted proteins and CAZymes during growth on each substrate (i.e. differences in transcription levels compared to the common maltose control for fair cross-comparisons). In total, we retrieved 5,536 genes from the six species and examined the transcription trends by grouping similarly-regulated genes using Self-Organizing Map (SOMs; see

Methods; Table S10). SOM clusters the genes according to the transcription profiles into nodes and locates nodes with similar patterns in the vicinity of each other on a map. The trained SOM containing 195 nodes was sliced into each of the substrate conditions. Such maps (i.e. Tatami maps) gave an overview of the transcriptomic patterns showing the mean normalized log2 fold change in each node (Fig. 6a). We observed that the genes up-regulated on the substrates were clustered in the bottom left area of the map. The overall sets of genes up-regulated in response to pine and aspen were similar, and the set of genes up-regulated in response to wheat straw globally overlapped those responsive to cellulose and to woody substrates. Looking at the distribution of CAZyme genes on the map for an overview of CAZyme expression regulations, we observed that almost all genes coding for CBM1-associated CAZymes were localized in nodes with up-regulation (Fig. 6b). We also noticed enrichment for glycoside hydrolase and carbohydrate esterase genes in the up-regulated fraction of the transcriptome. On the contrary, expansin and polysaccharide lyase genes (Fig. S9) were uniformly localized on the map and were not preferentially clustered in up-regulated nodes. Interestingly, the genes coding for glycosyltransferases were located outside the up-regulated nodes, indicating these genes were not differentially regulated by the substrates. Globally, the AA genes were evenly distributed over the map, except for AA9 LPMOs. In particular, AA9 LPMOs with a conserved C-terminal module (called X282; Lenfant *et al.*, 2017) were all clustered with cellulose responsive genes. For each fungus, one single AA9-X282 gene was identified, which was strongly up-regulated (28 to 500-fold change in transcript read counts) after three day-growth on cellulose (Table S11).

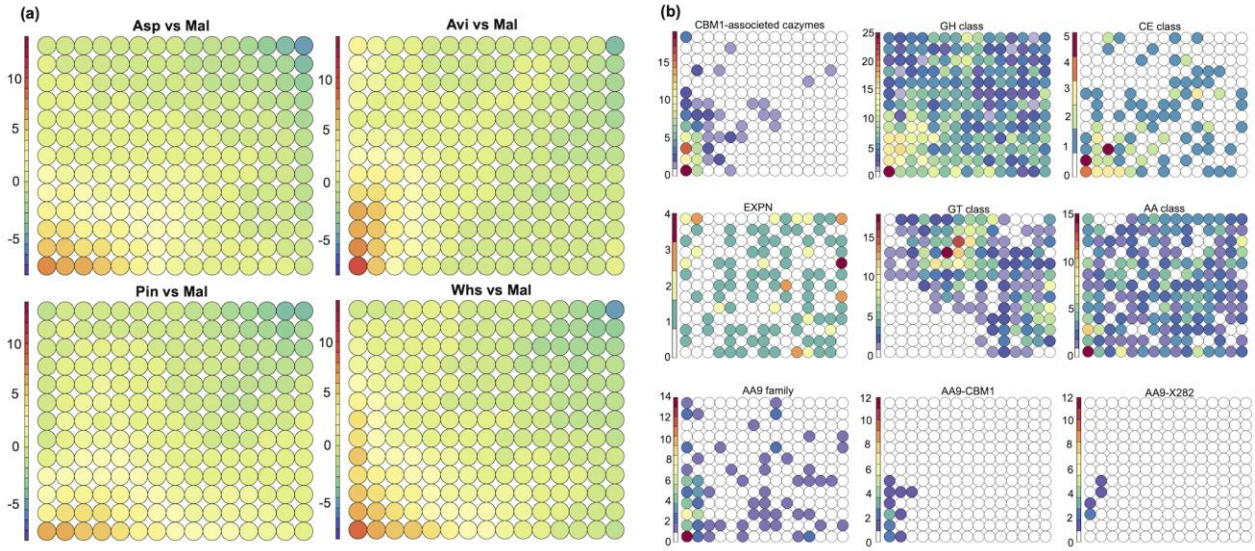


Figure 6 Overview of the transcription regulation of genes coding for CAZymes and predicted secreted proteins from 6 Polyporales species. The Tatami maps show, for each node, the mean log2 differential transcription of genes at day 3 on cellulose, aspen, pine, or wheat straw in comparison to maltose (a) and the counts for selected CAZyme genes in each node (b).

For a closer inspection of the conserved responses across the six species, the genes up-regulated on the tested substrates were retrieved from the nodes with the mean fold change in normalized transcript read counts > 4 after growth on cellulose, aspen, pine, or wheat straw in comparison to maltose. We first analyzed the expression of CAZyme genes. Out of the 63 CAZyme families analyzed, 11 were up-regulated in the six species in response to the four substrates (Fig. 7a; Fig. S10). This set of CAZyme hence represented a core response to (ligno)cellulose and contained the expected GH1 □□glucosidases, GH5_5 □-1,4-endoglucanasse, GH6 and GH7 cellobiohydrolases, GH131 glucanases, GH10 xylanases, CE1 and CE16 carbohydrate esterases, GH28 polygalacturonases and numerous oxidative AA9 LPMOs active on crystalline cellulose. Cellobiose dehydrogenases (AA8-AA3_1 modular enzymes), □□mannosidases (GH5_7) and □□glucuronidases that target glycosylated proteins (GH79) were found up-regulated in five out of the six species. We then examined the CAZyme genes specifically regulated in response to complex lignin and hemicellulose containing substrates, not to cellulose. We observed no conservation in the specific responses to complex substrates across the tested Polyporales (Fig. S11). These results suggested that cellulose was the main inducer of the shared response to the substrates within Polyporales. However, we cannot rule out the possibility that the absence of conservation in CAZyme response to complex substrates could be due to the analysis of a single time-point in our study, with the observation window being too narrow to see common changes.

Using a similar approach with no *a priori* on the functions of the differentially regulated genes, we looked at the genes with the same InterPro description that were up-regulated in at least five species. We found SSP genes with a thaumatin domain (IPR001938) were up-regulated in all species, except in *Pycnoporus cinnabarinus* (Fig. 7b; Table S12). In plants, thaumatins and thaumatin-like proteins are involved in plant defense against pathogens. However, the newly identified proteins had no similarity with known proteins from the SwissProt database, and had four predicted disulfide bridges, instead of eight in plant thaumatins, suggesting they might have different functions (Table S13). We also found genes coding predicted secreted proteins with a "Carboxylesterase, type_B" domain (IPR002018) were commonly up-regulated on all the tested substrates except in *Leiotrametes* sp. grown on cellulose (Fig. 7b). Using BLASTP search on Fungi proteins in the SwissProt database, we identified these proteins as candidate lipases (Table S13). Finally, we identified genes coding for predicted "cupredoxin" proteins (IPR008972) were up-regulated in response to pine in all species except *T. ljubarskyi*. We identified the copper protein ARB_05732-1 from the ascomycete *Arthroderma benhamiae* as the closest homolog among Fungi proteins (Table S13). Similar to ARB_05732-1, the candidate cupredoxin proteins we identified carried a predicted signal peptide for secretion and a probable glycosylphosphatidylinositol (GPI)-anchor domain.

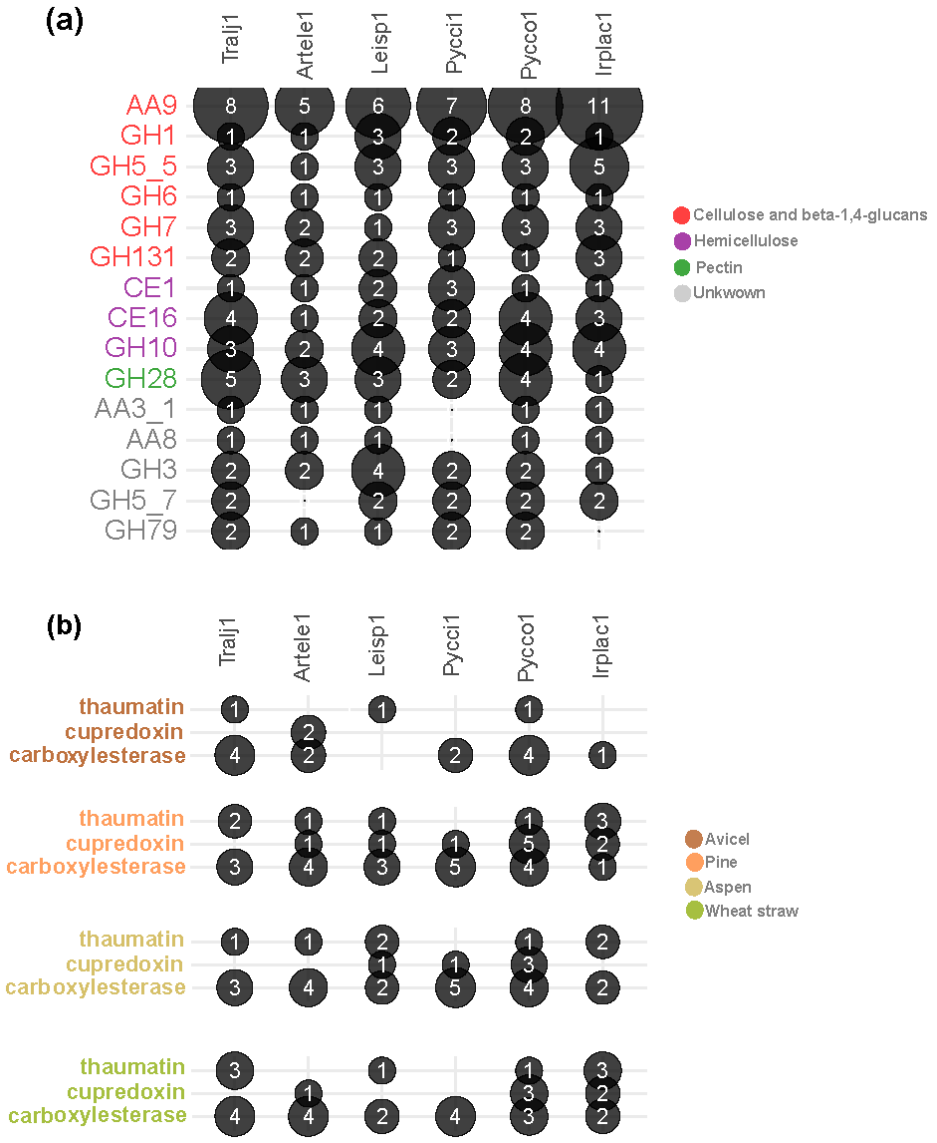


Figure 7 Count of CAZyme genes commonly up-regulated in response to cellulose, aspen, pine and wheat straw in at least 5 of the species. The colors indicate the substrates targeted by each CAZy family (a). Count of Carboxylesterase, cupredoxin and thaumatin genes up-regulated in response to cellulose (Avicel), pine, aspen and wheat straw (b).

Discussion

We found Polyporales might be around 183 million years old and have undergone intensive diversification, which is consistent with previous findings (Varga *et al.*, 2019). First, we observed poor synteny conservation within Polyporales clades, and the loss of synteny was not related to TE coverage. Further studies may be required to elucidate possible mechanisms for genome rearrangements. Second, on average 26% of the protein-coding genes were species-specific,

among which on average 81% were coding for predicted SSPs. The abundance of species-specific SSPs in Polyporales genomes is in line with previous observations on the genomes of saprotroph fungi including white-rot and brown-rot fungi, which encode numerous SSPs (Pellegrin *et al.*, 2015; Kim *et al.*, 2016), and suggests they might have a role in the determination of specific traits. For example, SSPs could be used as weapons or decoys in intra- and interspecific cell-to-cell communication in soil, wood, or the rhizosphere. They could also be used as regulators of quorum sensing (Feldman *et al.*, 2020). Our transcriptomic analysis on six Polyporales species revealed that about half of the SSPs with transcript evidence in each genome had no predicted function (55.6%, SD 4.4). And 8.7 to 32.7% (mean 22.9%, SD 9) of those were highly transcribed or up-regulated after three-days of growth on lignocellulosic substrates. Among the up-regulated SSPs with an InterPro domain, proteins with a thaumatin domain were up-regulated on pine and aspen in five out of the six species analyzed. Yet, no predicted function was identified for these SSPs by homology search. SSPs with a GPI membrane anchoring domain and a single cupredoxin (IPR008972) domain were induced in response to the lignocellulosic substrates in all six species analyzed. The cupredoxin fold allows copper-mediated electron transfer with the reduction of molecular oxygen to water. In eukaryotes, toxic copper (II) ions are reduced to copper (I) by monodomain cupredoxins before being transported inside the cell via the high-affinity copper transporters of the CTR family (Arnesano *et al.*, 2002). Further studies are required to elucidate whether SSPs harboring a cupredoxin domain are involved in copper homeostasis at the fungal cell wall, as recently suggested for LPMO-like copper containing proteins with a GPI-anchor domain located at the fungal cell wall of ectomycorrhizal and pathogenic fungi (Garcia-Santamarina *et al.*, 2020; Labourel *et al.*, 2020).

In a search for idiosyncrasies related to wood decay, we analyzed gene family expansion/contraction events inferred in early ancestors of the Polyporales. Surprisingly, we observed only few expansions for PCWDE in early Polyporales. White-rot fungi most likely evolved in the Carboniferous in the most recent common ancestor of the Auriculariales and higher Agaricomycetes (Floudas *et al.*, 2012; Nagy *et al.*, 2016), which predates the Polyporales MRCA by at least 100-150 million years (this study, Varga *et al.*, 2019). We explain the lack of clear PCWDE expansions in early Polyporales by the already diverse lignocellulosic toolkit present before the origin of the Polyporales (see e.g. Nagy *et al.*, 2016;), which may have provided early members of the Polyporales with the necessary tools to degrade plant cell walls. A remarkable exception of this pattern are class-II peroxidases (POD), that existed before, but further evolved for nonphenolic lignin degradation in the Polyporales (Ayuso-Fernández *et al.*, 2018), which coincided with duplication events also captured by our analyses. Among POD, MnP were identified in all white-rot Polyporales genomes, consistent with MnP being the first lignin-active POD type, which appeared in the common ancestor of white rotters (Floudas *et al.*, 2012; Nagy *et al.*, 2016). VP and LiP were present each in 69% of the white-rot Polyporales genomes

but had scarce occurrences in other white-rot Agaricomycetes we analyzed. Conversely, atypical MnP, with unconventional catalytic residues, were rare in Polyporales genomes (maximum two gene copies identified in three Polyporales species), whereas they are abundant in fungi from the Agaricomycetes orders (Ruiz-Dueñas et al., 2020). Whether these atypical MnP are involved in the degradation of lignin or lignin-derived polymers, such as the humic acids from forest litter is still to be determined.

Interestingly, our analysis of gene loss/gain within Polyporales showed different trajectories in the evolution of PCWDE gene families in the MRCA of the core polyporoid and phlebioid clade. It follows that functional differences in the wood-decay modes of white rotters from these two clades might also exist, in particular with regard to AA families, expansins and CBM1-containing proteins, but these will require further study.

We searched for transcriptomic evidence for shared mechanisms used by species from the core polyporoid and phlebioid clade as indicators of conserved toolkits of efficient wood decay. By comparing the transcriptomes of six species during growth on four different ligno(cellulosic) substrates, we found that genes coding for AA9 LPMOs appended to a CBM1 carbohydrate binding module were up-regulated in the six tested species. Our results are in line with a major role for AA9 LPMOs in cellulose degradation by Polyporales fungi (Berrin *et al.*, 2017) and the previous observation that the CBM1 module can enhance targeting of the LPMO catalytic module to the substrate (Bennati-Granier *et al.*, 2015; Chalak *et al.*, 2019). Interestingly, we observed the conserved up-regulation of genes coding for AA9 LPMOs with an atypical ~30 amino acid module fused to the C-terminal of the catalytic domain. The X282 module is to date only found in saprotroph Agaricomycetes and absent from the genome sequences of mycorrhizal Agaricomycetes (Fig. S12), and each of the six Polyporales genomes analyzed here contained one single AA9-X282 gene. In the case of the phlebioid *Irpex lacteus*, the enzyme also carried a CBM1 module. Further studies are required to understand the role of the X282 extension in the activity or specificity of the enzyme and its contribution to wood decay.

We also identified in the six analyzed species several lipase genes (triacylglycerol acylhydrolases, EC 3.1.1.3) strongly up-regulated during growth on lignocellulosic substrates. The secreted lipases could catalyze the hydrolysis of long chain triacylglycerol substrates (>C8) present in the wood extractives or at the fungal cell wall (Kadimaliev *et al.*, 2006), or could contribute to the degradation of suberin, a highly recalcitrant polyester embedded in the secondary cell walls in the phellem of tree barks and in the endodermis of roots (Martins *et al.*, 2014). Besides, wood lipid degradation generates free unsaturated fatty acids further converted by MnP-mediated peroxidation into lipid hydroperoxides which could be involved in lignin and xenobiotic degradation (Gutiérrez *et al.*, 2002).

Finally, we identified gene families with different expansion/contraction in early white-rot vs. brown-rot Polyporales. We identified gene expansions for peptidases, cytochrome p450s and transporters in early white-rot Polyporales. These findings support a role for these proteins in wood decay and are congruent with the previously observed up-regulation of these genes and the secretion of peptidases during growth on lignocellulosic substrates (e.g. this study; Hori *et al.*, 2014; Korripally *et al.*, 2015; Miyauchi *et al.*, 2017, 2018; Moudy & Stites, 2017). In early Polyporales we also observed gene expansion for several groups of regulator families, including F-box domain proteins, transcription factors, BTB/POZ proteins, RING/MYND type zinc finger proteins, LLR domain proteins and protein kinases. Although at the moment we do not have evidence for the involvement of these regulator families in wood decay, these gene duplication/loss patterns in the Polyporales is remarkable and lend support to the hypothesis that, while PCWDEs were in place in the ancestors of the Polyporales, the diversification of some regulatory families followed only after the origin of the order. This hypothesis is consistent with the generally low conservation of regulators of lignocellulose degradation (see e.g. Mattila *et al.*, 2020). Nevertheless, more detailed experimental follow-up are necessary to establish potential direct links with wood-decay.

Altogether, our results highlight Polyporales genomes as remarkable resources to unravel the enzymatic mechanisms of lignocellulose breakdown. Because they have evolved a diversity of enzymatic mechanisms to overcome the PCW polymer barrier, Polyporales are promising resources to inspire enzymatic toolkits aimed at ecofriendly processes to access PCW polysaccharides and to facilitate the extraction and bioconversion of plant molecules (Arantes *et al.*, 2012).

Materials and Methods

Sequencing, Assembly, and Annotation of 26 new Polyporales genomes

Genomic libraries were created using either “Illumina Regular Fragment, 300bp”, “Illumina Regular Long Mate Pair, 4kb, CLRS”, “PacBio Low Input 10kb”, or “PacBio > 10kb” protocols (Table S1, S2; see Supporting Information). PacBio SMRTbell templates were sequenced using the Pacific Biosciences RSII, or SEQUEL sequencers with the following parameters: Version C4 chemistry, 4h run times (RSII); Version 2.0 chemistry, 6h and 10h run times (SEQUEL). Libraries were assembled with Falcon (Maeder, 2018), finished with finisherSC (Lam *et al.*, 2015), and polished with Arrow (Maeder, 2019) or Quiver (Chin *et al.*, 2013). The Illumina genomic libraries

were sequenced with Illumina HiSeq2500 or HiSeq2000 using HiSeq TruSeq SBS sequencing kits, v4 (2x150 indexed run for “Regular Fragment” genomic libraries; 2x100 indexed run for “Regular LMP” genomic libraries). Illumina genomic libraries were assembled with AllPathsLG (Gnerre *et al.*, 2011) (Table S2). For RNA sequencing, the mycelia were ground in liquid nitrogen using a Freezer/Mill Cryogenic Grinder (SPEX Sample Prep, United Kingdom). Total RNA was extracted from 100 mg ground tissue using TRIZOL (Ambion), precipitated with isopropanol, resuspended in water and treated with RNase-Free DNase I (QIAGEN). Total RNA was precipitated again with LiCl and resuspended in DEPC-treated water as described in Miyauchi et al., 2020. RNA Sequencing was performed using an Illumina HiSeq 2000 or HiSeq 2500 instrument. All genomes were annotated using the JGI annotation pipeline (Grigoriev *et al.*, 2014). CAZymes and auxiliary activity enzymes (AA) were annotated as in Lombard *et al.*, 2014. Gene models from PODs were further classified based on the presence of known conserved residues i.e. manganese peroxidase (MnP), with a Mn(II)-oxidation site formed by three conserved acidic residues, lignin peroxidase (LiP) with an exposed catalytic tryptophan, and versatile peroxidase (VP) with both the Mn(II) oxidation site found in MnP and the catalytic tryptophan found in LiP (Floudas *et al.*, 2012; Ruiz-Dueñas *et al.*, 2013).

Reconstruction of the duplication/loss history of gene families

Whole proteomes of the 50 Polyporales species, and three from the orders Gloeophyllum and Corticiales (Table S3) were clustered to identify single-copy orthogroups. After removal of poorly aligned regions, the trimmed alignments of single-copy clusters were concatenated into a supermatrix to build a maximum likelihood tree that was used as input to MCMCTree (Puttick, 2019) for molecular clock dating analyses (see Sup. Information). To investigate the evolutionary history of each gene family, protein alignments of each cluster containing at least four proteins were first trimmed using trimAl 1.4 (Capella-Gutiérrez *et al.*, 2009) with a parameter –gt 0.2. Then we inferred a maximum likelihood phylogeny for each gene cluster and calculated Shimodaira-Hasegawa-like (SH-like) branch support values using the PTHREADS version of RAxML 8.2.12 under the PROTGAMMAWAG model. Next, we midpoint-rooted and reconciled the rooted gene trees with the species tree using Notung 2.9 (Chen *et al.*, 2000) with an edge-weight threshold of 0.95. Gene trees were then further processed by cutting them at basal duplication nodes where possible, which resulted in a final number of 36,153 clusters containing at least four proteins. Together with the clusters containing less than four proteins, we reconstructed the duplication/loss history of 288,253 protein clusters across the species tree using the COMPARE pipeline (Nagy *et al.*, 2014, 2016). Duplication/loss rates were computed by dividing the number of inferred duplication/loss events by the length of the respective branch of the species tree the event occurred on. Gene families were functionally characterized based on

InterPro annotations. Graphical maps of gene duplication/loss histories were generated using custom scripts in R.

Protein conservation across Polyporales

We downloaded the protein sequences from the JGI MycoCosm database, for the 50 Polyporales species. A reciprocal protein BLAST was performed using BLAST+2.7.1. Orthologous genes were then clustered using FASTORTHO (Wattam *et al.*, 2014) with 50% identity and 50% length coverage thresholds. The predicted proteins shared by all 50 genomes were counted as core proteins. The ones shared by at least two genomes were counted as dispensable. The sequences of species-specific proteins were further used to search for similarities in all available fungal proteomes by running BLAST on the MycoCosm (50% identity and 50 % coverage).

Identification of transposable elements

We identified transposable elements (TE) using a previously described method (Payen *et al.*, 2016). Briefly, REPEATSCOUT 1.0.5 (Price *et al.*, 2005) with default parameters was used to generate *de novo* predictions of repeat elements in the unmasked genomes. Repeated sequences found more than ten times in the genomes were annotated using TBLASTX (Altschul *et al.*, 1990) against the fungal reference sequences from REPBASE version 22.08 (http://www.girinst.org/server/RepBase/index.php). The coverage of TEs in the genomes, including unknown TEs, was estimated by REPEATMASKER open 4.0.6 (http://www.repeatmasker.org). The results were integrated and visualized using the Transposon Identification Nominative Genome Overview custom script in R (TINGO; Morin *et al.*, 2019).

Genome macrosynteny

The analysis of genome macrosynteny across Polyporales was done on 23 selected genomes that had their ten largest scaffolds covering more than 40% of their genomes. Pair-wise comparisons and identification of syntenic blocks were done on the ten largest scaffolds using the R package DECIPHER (Wright, 2015), with default parameters and without masking repeat sequences. The macrosynteny between two species was visualized with the R package Circlize (Gu *et al.*, 2014). Data manipulation, integration, and visualization were coordinated using a set of custom R scripts Syntenic Governance Overview (SynGO, available on request).

Transcriptomics of six Polyporales species

Artolenzites elegans BRFM 1663, *Leiotrametes* sp. BRFM 1775, *Pycnoporus cinnabarinus* BRFM 137, *Pycnoporus coccineus* BRFM 310, *Trametes ljubarskyi* BRFM 1659 and *Irpex lacteus* CCBAS Fr. 238 617/93 were grown for three days in liquid culture medium supplemented with cellulose Avicel PH 101 (Fluka) (15 $g.L^{-1}$), ground and sifted wheat straw fragments < 2 mm (15 $g.L^{-1}$), ground and sifted *Pinus halepensis* wood fragments < 2 mm (15 $g.L^{-1}$) or 1 mm Wiley-milled *Populus tremuloides* (15 $g.L^{-1}$) (see Sup. Information). For control, each strain was grown in the same medium supplemented with maltose (20 $g.L^{-1}$). Each culture was done in triplicate. RNA sequencing was performed as described above. The sequence reads were trimmed for quality and aligned to the reference genome to generate the raw gene counts (see Sup. Information). Raw gene counts were used to evaluate the level of correlation between biological replicates using Pearson's correlation. DESeq2 (version 1.2.10; Love *et al.*, 2014) was subsequently used to determine which genes were differentially expressed between pairs of conditions. The parameters used to identify differentially expressed genes were Bonferroni-adjusted p-value <0.05 and Benjamini–Hochberg-adjusted p-value <0.05. The transcript log2 fold changes were transformed into the range zero to one using unity-based normalization (Aksoy & Haralick, 2001; Fig. S13). The normalized values were used to train the Self-organizing map (SOM) with 975,000 training iterations (195 map units x 5000 times), using the visual multi-omics pipeline Self-organizing map Harbouring Informative Nodes with Gene Ontology (SHIN+GO; Miyauchi *et al.*, 2017, 2020). The Tatami maps generated from SHIN+GO represented clustered genes from the six strains with similar transcription profiles.

Comparison of the predicted secretomes

The secreted proteins were predicted using a custom pipeline described in Pellegrin *et al.*, 2015 (Supp. Information). Predicted secreted proteins smaller than 300 amino acids were considered as small secreted proteins (SSPs). CAZymes were annotated as in Lombard *et al.*, 2014. The impacts of phylogeny and lifestyle on secreted CAZyme counts were assessed using the function adonis2 from the Vegan R package (Oksanen *et al.*, 2012) after converting the Polyporales species tree into evolutionary distance using R package *ape* (Paradis *et al.*, 2004) and CAZyme counts into distance matrix using Vegdist function. RVAideMemoire package (Maxim, 2018) was used to check significant difference among Polyporales clades using a pair-wise PERMANOVA (Mesny, 2020).

Data availability

The assemblies and annotations of 26 newly sequenced genomes are available at the US Department of Energy Joint Genome Institute (JGI) fungal genome portal

MycoCosm (https://mycocosm.jgi.doe.gov; Grigoriev *et al.*, 2014) and in the DDBJ/EMBL/GenBank repository (https://www.ncbi.nlm.nih.gov; Table S1). The RNASeq data generated in the current study are available at Gene Expression Omnibus repository under accession nos. GSE82427, GSE82419, GSE156901.

Acknowledgments

This work was supported by the U.S. Department of Energy Joint Genome Institute, a Department of Energy Office of Science User Facility (grant # DE-AC02-05CH11231, DE-SC0019427 to I.V.G. and B.M.); Institut Carnot 3BCAR, the French National Research Institute for Agriculture, Food and Environment, The Region Provence Alpes Côte d'Azur and the Groupement de Recherche Génomique Environnementale to H.H.; the Laboratory of Excellence ARBRE (grant # ANR-11-LABX-0002-01 to F.M.); the Region Lorraine and the European Regional Development Fund to F.M.; the Hungarian Academy of Sciences' Momentum Program (grant # LP2019-13/2019 to L.G.N.); the Spanish Ministry of Economy, Industry and Competitiveness (grant # BIO2017-86559-R to A.T.M.); the Consejo Superior de Investigaciones Científicas (grant # PIE-201620E081 to A.T.M.); the Agencia Estatal de Investigación, the European Regional Development Fund and the Ministry of Science, Innovation and Universities (grant # RTI2018-093683-B-I00 to L.D.E. and M.J.M.) and the Czech Science Foundation (grant # 17-20110S to P.B. and M.Š.).

Conflict of interest

The authors have no conflict of interest.

Author contributions

LLM, MNR, FM conceived the work. IVG and KB coordinated JGI projects. AF, DC, DN, SG, SMa, LLM, MN, ATM, LdE, MJM, CN, PB, MS, SM, JM, JWS produced the biological material. JP, WA, KLB, HH, HN, AK, KB, AL, RR, SA, BH, ED, SM generated and analyzed the data. LGN, MV, ZM, BB performed the phylogenomics analyses. SM did the SOM analysis. HH and SM did the comparative genomics analysis. HH, SM, LGN, and MNR wrote the manuscript with the help of all co-authors.

References

Aksoy S, Haralick RM. **2001**. Feature normalization and likelihood-based similarity measures for image retrieval. *Pattern Recognition Letters* **22**: 563–582.

Arantes V, Jellison J, Goodell B. **2012**. Peculiarities of brown-rot fungi and biochemical Fenton reaction with regard to their potential as a model for bioprocessing biomass. *Applied microbiology and biotechnology* **94**: 323–338.

Arnesano F, Banci L, Bertini I, Thompsett AR. **2002**. Solution structure of CopC: A cupredoxin-like protein involved in copper homeostasis. *Structure* **10**: 1337–1347.

Ayuso-Fernández I, Ruiz-Dueñas FJ, Martínez AT. **2018**. Evolutionary convergence in lignin-degrading enzymes. *Proceedings of the National Academy of Sciences* **115**: 6428–6433.

Bennati-Granier C, Garajova S, Champion C, Grisel S, Haon M, Zhou S, Fanuel M, Ropartz D, Rogniaux H, Gimbert I, *et al.* 2015. Substrate specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by *Podospora anserina*. *Biotechnology for Biofuels* **8**: 90.

Berrin J-G, Navarro D, Couturier M, Olivé C, Grisel S, Haon M, Taussac S, Lechat C, Courtecuisse R, Favel A, *et al.* 2012. Exploring the natural fungal biodiversity of tropical and temperate forests toward improvement of biomass conversion. *Applied and environmental microbiology* **78**: 6483–6490.

Berrin J-G, Rosso M-N, Abou Hachem M. **2017**. Fungal secretomics to probe the biological functions of lytic polysaccharide monooxygenases. *Carbohydrate research* **448**: 155–160.

Blanchette RA. **1984**. Screening wood decayed by white rot fungi for preferential lignin degradation. *Applied and Environmental Microbiology* **48**: 647–653.

Blanchette, R. A., Krueger, E. W., Haight, J. E., Akhtar, M., & Akin, D. E. **1997**. Cell wall alterations in loblolly pine wood decayed by the white-rot fungus, *Ceriporiopsis subvermispora*. *Journal of Biotechnology*, 53(2-3), 203-213.

Boberg JB, Finlay RD, Stenlid J, Ekblad A, Lindahl BD. **2014**. Nitrogen and carbon reallocation in fungal mycelia during decomposition of boreal forest litter (RA Wilson, Ed.). *PLoS ONE* **9**: e92897.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. **2009**. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England)* **25**: 1972–1973.

Chalak A, Villares A, Moreau C, Haon M, Grisel S, d'Orlando A, Herpoël-Gimbert I, Labourel A, Cathala B, Berrin J-G. **2019**. Influence of the carbohydrate-binding module on the activity of a fungal AA9 lytic polysaccharide monooxygenase on cellulosic substrates. *Biotechnology for Biofuels* **12**: 206.

Chen K, Durand D, Farach-Colton M. **2000**. NOTUNG: A program for dating gene duplications and optimizing gene family trees. *Journal of Computational Biology* **7**: 429–447.

Chen J, Heikkinen J, Hobbie EA, Rinne-Garmston KT, Penttilä R, Mäkipää R. **2019**. Strategies of carbon and nitrogen acquisition by saprotrophic and ectomycorrhizal fungi in Finnish boreal Picea abies-dominated forests. *Fungal Biology* **123**: 456–464.

Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, *et al.* 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods* **10**: 563–569.

Couturier M, Ladevèze S, Sulzenbacher G, Ciano L, Fanuel M, Moreau C, Villares A, Cathala B, Chaspoul F, Frandsen KE, *et al.* 2018. Lytic xylan oxidases from wood-decay fungi unlock biomass

degradation. *Nature chemical biology* **14**: 306–310.

Deroy A, Saiag F, Kebbi-Benkeder Z, Touahri N, Hecker A, Morel-Rouhier M, Colin F, Dumarcay S, Gérardin P, Gelhaye E. **2015**. The GSTome reflects the chemical environment of white-rot fungi. *PloS one* **10**: e0137083.

Eastwood DC, Floudas D, Binder M, Majcherczyk A, Schneider P, Aerts A, Asiegbu FO, Baker SE, Barry K, Bendiksby M, *et al.* 2011. The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. *Science (New York, N.Y.)* **333**: 762–765.

Feldman D, Yarden O, Hadar Y. **2020**. Seeking the roles for fungal small-secreted proteins in affecting saprophytic lifestyles. *Frontiers in Microbiology* **11**: 455.

Fernandez-Fueyo E, Ruiz-Dueñas FJ, Ferreira P, Floudas D, Hibbett DS, Canessa P, Larrondo LF, James TY, Seelenfreund D, Lobos S, *et al.* 2012. Comparative genomics of *Ceriporiopsis subvermispora* and *Phanerochaete chrysosporium* provide insight into selective ligninolysis. *Proc Natl Acad Sci U S A.* **109**: 5458–5463.

Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martínez AT, Otillar R, Spatafora JW, Yadav JS, *et al.* 2012. The paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* **336**: 1715–1719.

Garcia-Santamarina S, Probst C, Festa RA, Ding C, Smith AD, Conklin SE, Brander S, Kinch LN, Grishin N V, Franz KJ, *et al.* 2020. A lytic polysaccharide monooxygenase-like protein functions in fungal copper import and meningitis. *Nature Chemical Biology* **16**: 337–344.

Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, *et al.* 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A.* **108**: 1513–1518.

Grigoriev I V, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, *et al.* 2014. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* **42**: D699–D704.

Gu Z, Gu L, Eils R, Schlesner M, Brors B. **2014**. Circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**: 2811–2812.

Gutiérrez A, del Río JC, Martínez-Íñigo MJ, Martínez MJ, Martínez ÁT. **2002**. Production of new unsaturated lipids during wood decay by ligninolytic basidiomycetes. *Applied and Environmental Microbiology* **68**: 1344–1350.

Hori C, Gaskell J, Igarashi K, Kersten P, Mozuch M, Samejima M, Cullen D. **2014**. Temporal alterations in the secretome of the selective ligninolytic *fungus Ceriporiopsis subvermispora* during growth on aspen wood reveal this organism's strategy for degrading lignocellulose. *Appl Environ Microbiol.* **80**: 2062–2070.

Justo A, Miettinen O, Floudas D, Ortiz-Santana B, Sjökvist E, Lindner D, Nakasone K, Niemelä T, Larsson K-H, Ryvarden L, *et al.* 2017. A revised family-level classification of the Polyporales (Basidiomycota). *Fungal biology* **121**: 798–824.

Kadimaliev DA, Nadezhina OS, Atykyan NA, Revin V V, Samuilov VD. **2006**. Interrelation between the composition of lipids and their peroxidation products and the secretion of ligninolytic enzymes during growth of *Lentinus* (*Panus*) *tigrinus*. *Microbiology* **75**: 563–567.

Kim K-T, Jeon J, Choi J, Cheong K, Song H, Choi G, Kang S, Lee Y-H. **2016**. Kingdom-wide analysis of fungal small secreted proteins (SSPs) reveals their potential role in host association. *Frontiers in Plant Science* **7**: 186.

Kirk P, Cannon P, Minter D, Stalpers J. **2009**. Dictionary of the Fungi. *Mycol Res* **113**: 908–910.

Knobloch E, Kotlaba F. **1994**. *Trametites eocenicus*, a new fossil polypore from the Bohemian Eocene. *Czech Mycol* **47**: 207–214.

Korripally P, Hunt CG, Houtman CJ, Jones DC, Kitin PJ, Cullen D, Hammel KE. **2015**. Regulation of gene expression during the onset of ligninolytic oxidation by *Phanerochaete chrysosporium* on spruce wood. *Applied and Environmental Microbiology* **81**: 7802 –7812.

Kuroyama H, Tsutsui N, Hashimoto Y, and Tsumuraya Y. 2001. Purification and characterization of a beta-glucuronidase from *Aspergillus niger*. *Carbohydr Res*. 333, 27-39.

Kuuskeri J, Häkkinen M, Laine P, Smolander O-P, Tamene F, Miettinen S, Nousiainen P, Kemell M, Auvinen P, Lundell T. **2016**. Time-scale dynamics of proteome and transcriptome of the white-rot fungus *Phlebia radiata*: growth on spruce wood and decay effect on lignocellulose. *Biotechnology for Biofuels* **9**: 192.

Labourel A, Frandsen KEH, Zhang F, Brouilly N, Grisel S, Haon M, Ciano L, Ropartz D, Fanuel M, Martin F, *et al.* 2020. A fungal family of lytic polysaccharide monooxygenase-like copper proteins. *Nature Chemical Biology* **16**: 345–350.

Lam K-K, LaButti K, Khalak A, Tse D. **2015**. FinisherSC: a repeat-aware tool for upgrading de novo assembly using long reads. *Bioinformatics* **31**: 3207–3209.

Lenfant N, Hainaut M, Terrapon N, Drula E, Lombard V, Henrissat B. **2017**. A bioinformatics analysis of 3400 lytic polysaccharide oxidases from family AA9. *Carbohydrate Research* **448**: 166–174.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. **2014**. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**: D490-5.

Love MI, Huber W, Anders S. **2014**. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**: 550.

Lundell TK, Mäkelä MR, de Vries RP, Hildén KS. **2014**. Chapter Eleven - Genomics, Lifestyles and Future Prospects of Wood-Decay and Litter-Decomposing Basidiomycota. In: F M Martin (ed.), Advances in Botanical Research : Fungi . vol. 70 , Academic Press , London , pp. 329-370 .

Martin R, Gazis R, Skaltsas D, Chaverri P, Hibbett D. **2015**. Unexpected diversity of basidiomycetous endophytes in sapwood and leaves of Hevea. *Mycologia* **107**: 284–297.

Martínez AT, Ruiz-Dueñas FJ, Camarero S, Serrano A, Linde D, Lund H, Vind J, Tovborg M, Herold-Majumdar OM, Hofrichter M, *et al.* 2017. Oxidoreductases on their way to industrial biotransformations. *Biotechnology advances* **35**: 815–831.

Martínez AT, Speranza M, Ruiz-Dueñas FJ, Ferreira P, Camarero S, Guillén F, Martínez MJ, Gutiérrez A, del Río JC. **2005**. Biodegradation of lignocellulosics: microbial, chemical, and enzymatic aspects of the fungal attack of lignin. *International microbiology : the official journal of the Spanish Society for Microbiology* **8**: 195–204.

Martins I, Hartmann DO, Alves PC, Martins C, Garcia H, Leclercq CC, Ferreira R, He J, Renaut J, Becker JD, *et al.* 2014. Elucidating how the saprophytic fungus *Aspergillus nidulans* uses the plant polyester suberin as carbon source. *BMC Genomics* **15**: 613.

Maeder H. 2018. FALCON: experimental PacBio diploid assembler. [WWW document] URL https://github.com/PacificBiosciences/FALCON.

Maeder H. 2019. PacBio® variant and consensus caller. [WWW document] URL https://github.com/PacificBiosciences/GenomicConsensus

Mattila HK, Mäkinen M, Lundell T. **2020**. Hypoxia is regulating enzymatic wood decomposition and intracellular carbohydrate metabolism in filamentous white rot fungus. *Biotechnology for Biofuels* **13**: 26.

Maxim H. **2018**. RVAideMemoire: testing and plotting procedures for biostatistics. R Package Version 0.9-69.

Mesny F. 2020. Detecting the effect of biological categories on genomes composition. [WWW document] URL https://github.com/fantin-mesny/Effect-Of-Biological-Categories-On-Genomes-Composition [accessed July 2020].

Miettinen, O., Riley, R., Barry, K., Cullen, D., de Vries, R. P., Hainaut, M., Hatakka, A., Henrissat, B., Hildén, K., Kuo, R., *et al*. 2016. Draft Genome Sequence of the White-Rot Fungus *Obba rivulosa* 3A-2. *Genome announcements*, 4(5), e00976-16.

Miyauchi S, Hage H, Drula E, Lesage-Meessen L, Berrin J-G, Navarro D, Favel A, Chaduli D, Grisel S, Haon M, *et al.* 2020. Conserved white-rot enzymatic mechanism for wood decay in the Basidiomycota genus *Pycnoporus*. *DNA research* **27**: *dsaa011*..

Miyauchi S, Navarro D, Grisel S, Chevret D, Berrin J-G, Rosso M-N. **2017**. The integrative omics of white-rot fungus *Pycnoporus coccineus* reveals co-regulated CAZymes for orchestrated lignocellulose breakdown. *PLOS ONE* **12**: e0175528–e0175528.

Miyauchi S, Rancon A, Drula E, Hage H, Chaduli D, Favel A, Grisel S, Henrissat B, Herpoël-Gimbert I, Ruiz-Dueñas FJ, *et al.* 2018. Integrative visual omics of the white-rot fungus *Polyporus brumalis* exposes the biotechnological potential of its oxidative enzymes for delignifying raw plant biomass. *Biotechnology for biofuels* **11**: 201.

Morin E, Miyauchi S, San Clemente H, Chen ECH, Pelin A, de la Providencia I, Ndikumana S, Beaudet D, Hainaut M, Drula E, *et al.* 2019. Comparative genomics of *Rhizophagus irregularis*, *R. cerebriforme*, *R. diaphanus* and *Gigaspora rosea* highlights specific genetic features in Glomeromycotina. *New Phytologist* **222**: 1584–1598.

Moudy M, Stites W. **2017**. Investigation of methionine sulfoxide formation as a regulator of proteolysis. *Biophysical Journal* **112**: 202a.

Nagy LG, Ohm RA, Kovács GM, Floudas D, Riley R, Gácser A, Sipiczki M, Davis JM, Doty SL, de Hoog GS, *et al.* 2014. Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. *Nature Communications* **5**: 4471.

Nagy LG, Riley R, Bergmann PJ, Krizsán K, Martin FM, Grigoriev I V, Cullen D, Hibbett DS. **2016**. Genetic bases of fungal white rot wood decay predicted by phylogenomic analysis of correlated gene-phenotype evolution. *Molecular Biology and Evolution* **34**: 35–44.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin P, O'Hara RB, Simpson G, Solymos P, Stevenes MHH, Wagner H. **2012**. Vegan: Community Ecology Package. R package version 2.0-2.

Paradis E, Claude J, Strimmer K. **2004**. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**: 289–290.

Payen T, Murat C, Martin F. **2016**. Reconstructing the evolutionary history of gypsy retrotransposons in the Périgord black truffle (*Tuber melanosporum* Vittad.). *Mycorrhiza* **26**: 553–563.

Pellegrin C, Morin E, Martin FM, Veneault-Fourrey C. **2015**. Comparative analysis of secretomes from ectomycorrhizal fungi with an emphasis on small-secreted proteins. *Frontiers in Microbiology* **6**: 1278.

Presley GN, Schilling JS. **2017**. Distinct growth and secretome strategies for two taxonomically divergent brown rot fungi. *Applied and environmental microbiology* **83**: e02987-16.

Price AL, Jones NC, Pevzner PA. **2005**. De novo identification of repeat families in large genomes. *Bioinformatics* **21 Suppl 1**: i351-8.

Puttick MN. **2019**. MCMCtreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. *Bioinformatics* **35**: 5321–5322.

Qhanya LB, Matowane G, Chen W, Sun Y, Letsimo EM, Parvez M, Yu J-H, Mashele SS, Syed K. **2015**. Genome-wide annotation and comparative analysis of cytochrome P450 monooxygenases in Basidiomycete biotrophic plant pathogens. *PLOS ONE* **10**: e0142100.

Rayner ADM, Boddy L, Dowson CG. **1987**. Temporary parasitism of *Coriolus* spp. by *Lenzites betulina*: A strategy for domain capture in wood decay fungi. *FEMS Microbiology Letters* **45**: 53–58.

Riley R, Salamov AA, Brown DW, Nagy LG, Floudas D, Held BW, Levasseur A, Lombard V, Morin E, Otillar R, *et al.* 2014. Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *Proceedings of the National Academy of Sciences* **111**: 9923–9928.

Rivera-Hoyos CM, Morales-Álvarez ED, Poutou-Piñales RA, Pedroza-Rodríguez AM, RodrÍguez-Vázquez R, Delgado-Boada JM. **2013**. Fungal laccases. *Fungal Biology Reviews* **27**: 67–82.

Ruiz-Dueñas FJ, Lundell T, Floudas D, Nagy LG, Barrasa JM, Hibbett DS, Martínez AT. **2013**. Lignin-degrading peroxidases in Polyporales: an evolutionary survey based on 10 sequenced genomes. *Mycologia* **105**: 1428–1444.

Ruiz-Dueñas, F. J., Barrasa, J. M., Sánchez-García, M., Camarero, S., Miyauchi, S., Serrano, A., et al. 2020. Genomic Analysis Enlightens Agaricales Lifestyle Evolution and Increasing Peroxidase Diversity. *Molecular Biology and Evolution*.

Schilling JS, Kaffenberger JT, Held BW, Ortiz R, Blanchette RA. **2020**. Using wood rot phenotypes to illuminate the "gray" among decomposer Fungi. *Frontiers in Microbiology* **11**: 1288.

Sokolski S, Bernier-Cardou M, Piché Y, Bérubé JA. **2007**. Black spruce (*Picea mariana*) foliage hosts numerous and potentially endemic fungal endophytes. *Canadian Journal of Forest Research* **37**: 1737–1747.

Song J, Cui B-K. **2017**. Phylogeny, divergence time and historical biogeography of *Laetiporus* (Basidiomycota, Polyporales). *BMC evolutionary biology* **17**: 102.

Suzuki H, MacDonald J, Syed K, Salamov A, Hori C, Aerts A, Henrissat B, Wiebenga A, VanKuyk PA, Barry K, *et al.* 2012. Comparative genomics of the white-rot fungi, *Phanerochaete carnosa* and *P. chrysosporium*, to elucidate the genetic basis of the distinct wood types they colonize. *BMC genomics* **13**: 444.

Syed K, Shale K, Pagadala NS, Tuszynski J. **2014**. Systematic identification and evolutionary analysis of catalytically versatile cytochrome p450 monooxygenase families enriched in model basidiomycete fungi. *PLOS ONE* **9**: e86683.

Varga T, Krizsán K, Földi C, Dima B, Sánchez-García M, Sánchez-Ramírez S, Szöllősi GJ, Szarkándi JG, Papp V, Albert L, *et al.* 2019. Megaphylogeny resolves global patterns of mushroom evolution. *Nature ecology & evolution* **3**: 668–678.

Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, *et al.* 2014. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Research* **42**: D581-91.

Wright ES. **2015**. DECIPHER: harnessing local sequence context to improve protein multiple sequence

alignment. *BMC Bioinformatics* **16**: 322.

Wu B, Gaskell J, Held BW, Toapanta C, Vuong T, Ahrendt S, Lipzen A, Zhang J, Schilling JS, Master E, *et al.* 2018. Substrate-specific differential gene expression and RNA editing in the brown rot fungus *Fomitopsis pinicola*. *Applied and Environmental Microbiology* **84**: e00991-18.

Zhang J, Silverstein KAT, Castaño JD, Figueroa M, Schilling JS. **2019**. Gene regulation shifts shed light on fungal adaption in plant biomass decomposers. *mBio* **10**: e02176-19.

Zhou S, Raouche S, Grisel S, Navarro D, Sigoillot J-C, Herpoël-Gimbert I. **2015**. Solid-state fermentation in multi-well plates to assess pretreatment efficiency of rot fungi on lignocellulose biomass. *Microbial biotechnology* **8**: 940–949.

Supporting Information

Methods S1 Construction of genomic libraries for 26 Polyporales genomes

Genomic libraries were created using either "Illumina Regular Fragment, 300bp", "Illumina Regular LMP, 4kb, CLRS", "PacBio Low Input 10kb", or "PacBio > 10kb" protocols (Table S2). For "Illumina Regular Fragment" libraries, 100 ng DNA was sheared and treated with end repair, A-tailing, and ligation of Illumina compatible adapters (IDT, Inc) using the KAPA-Illumina library creation kit (KAPA biosystems). For "Regular LMP" libraries, 5 µg DNA was sheared and DNA fragments were treated with end repair, ligated with biotinylated adapters containing loxP, then circularized via recombination by a Cre excision reaction (NEB), randomly sheared, and finally treated with end repair and A-tailing using the KAPA-Illumina library creation kit followed by immobilization of mate pair fragments on streptavidin beads (Invitrogen). Illumina compatible adapters (IDT, Inc) were ligated to the mate pair fragments and 8 PCR cycles were used to enrich the final library (KAPA Biosystems). For "PacBio 10kb" or "PacBio >10kb" genomic libraries, the DNA fragments obtained from 1 µg or 5 µg of gDNA were prepared using the PacBio SMRTbell template preparation kit and ligated to PacBio hairpin adapters. The resulting SMRTbell templates were purified using exonuclease treatments and size-selected using AMPure PB beads.

Methods S2 Construction and sequencing of transcript libraries for 26 Polyporales species

Illumina transcriptomic libraries were prepared using plate- or tube-based protocols. Plate-based RNA sample preparation was performed on the PerkinElmer Sciclone NGS robotic liquid handling system using Illumina TruSeq Stranded mRNA HT sample prep kit utilizing poly-A selection of mRNA following the standard protocol provided by the manufacturer, and with the following conditions: 1 µg per sample total starting material, and 8 PCR cycles. Tube-based cDNA libraries were prepared using the Illumina Truseq Stranded RNA LT kit. mRNA was purified from 1 µg of total RNA using magnetic beads containing poly-T oligos, then fragmented and reverse transcribed using random hexamers and SSII (Invitrogen) followed by second-strand synthesis. The fragmented cDNA was treated with end-pair, A-tailing, adapter ligation, and 8 PCR cycles. The transcriptomic libraries were sequenced with Illumina HiSeq2500 or HiSeq2000 using HiSeq TruSeq SBS sequencing kits, v4 (2x150 indexed run). Illumina transcriptomic libraries were assembled with Trinity (Grabherr *et al.*, 2011) or Rnnotator (Martin *et al.*, 2010).

Methods S3 Predicted secretomes

The prediction of secreted proteins in the 50 Polyporales genomes was performed using a custom bioinformatic pipeline described in (Pellegrin *et al.*, 2015). Briefly, SignalP v.4.1 (Petersen *et al.*, 2011) was used as a first filter to select proteins containing a signal peptide, then TMHMM v.2.0 with default parameters was used to filter out those with additional

transmembrane helix. Among the remaining proteins, the ones with subcellular localizations assigned as a secretory pathway with TargetP v.1.1 (Emanuelsson *et al.*, 2000) or as extracellular with WolfPsort v.2.0 using the option "fungi" (Horton *et al.*, 2007) were further selected as predicted secreted. PS-SCAN v1.8.8 was finally used to exclude proteins having the KDEL motif (Lys-Asp-Glu-Leu), responsible for the retention of the proteins in the endoplasmic reticulum.

Methods S4 Phylogeny of Agaricomycetes and molecular clock analysis

Whole proteomes of the 50 Polyporales species and 57 additional Agaricomycetes species (Table S3) were compared using all-vs-all mpiBLAST 1.6.0 (Darling *et al.*, 2003) with 50% bidirectional coverage and an e-value cutoff of 10-5. Proteins were then clustered using the MCL algorithm (Stichting *et al.*, 2000), with an inflation parameter 2.0. To infer a maximum likelihood phylogeny for the 107 species, we first filtered for single copy orthogroups with at least 70% taxon occupancy. This resulted in 540 clusters that were aligned using mafft 7.4.07 (Katoh & Standley, 2013) with the --Li-NSI method, except for clusters 1, 2, 3, 8, 13, 13, 19, 21, 38, where the --auto method was used due to computational limitations. Poorly aligned regions were removed using trimAl 1.4 (Capella-Gutiérrez *et al.*, 2009) with the parameter –gt 0.9. Trimmed alignments of single copy clusters were concatenated into a supermatrix, from which we inferred a species tree for the 107 species using the PTHREADS version of RAxML 8.2.12 (Stamatakis, 2014) under the PROTGAMMAWAG model. The model was partitioned by gene and bootstrapping was performed on the dataset in 100 replicates. The maximum likelihood tree along with the concatenated matrix were used as input to MCMCTree (Puttick, 2019) for molecular clock dating analyses. We used six fossils (*Appianoporites vancouverensis*, *Ganodermites lybicus*, *Trametites eocenicus*, *Archaemarasmius legettii*, *Palaeoagaricites antiquus* and a Suilloid ECM fossil; see Varga *et al.*, 2019), and *Ganodermites lybicus* (Fleischmann *et al.*, 2007), which provides a shallow calibration point within the Polyporales.

Methods S5 Culture conditions

Artolenzites elegans BRFM 1663, *Leiotrametes* sp. BRFM 1775, *Pycnoporus cinnabarinus* BRFM 137, *Pycnoporus coccineus* BRFM 310, *Trametes ljubarskyi* BRFM 1659 and *Irpex lacteus* CCBAS Fr. 238 617/93 were grown in liquid cultures at 30°C in a rotary shaker at 120 rpm in 250-ml Erlenmeyer flasks containing 100 ml of culture medium diammonium tartrate (1.84 g.L-1), yeast extract (2.5 g.L-1), maltose (2.5 g.L-1), KH_2PO_4 (0.2 g.L-1), $CaCl_2$, $2H_2O$ (1.32.10-2 g.L-1), $MgSO_4$, $7H_2O$ (0.5 g.L-1), $FeSO_4$, 7H2O (0.07 g.L-1), $ZnSO_4$, $7H_2O$ (7.77.10-3 g.L-1), $MnSO_4$, H_2O (3.63.10-3 g.L-1), $CuSO_4$, $5H_2O$ (7.2.10-4 g.L-1) and thiamine (0.25 g.L-1) supplemented with cellulose Avicel PH 101 (Fluka) (15 g.L-1), ground and sifted wheat straw fragments < 2 mm (15 g.L-1), ground and sifted *Pinus halepensis* wood fragments < 2 mm (15 g.L-1) or 1 mm Wiley-milled *Populus tremuloides* (15 g.L-1). For control, each strain was grown in the same medium supplemented with maltose (20 g.L-1). Each culture was done in triplicate. Inocula of the liquid

cultivations were prepared as described in (Herpoël *et al.*, 2000).

Methods S6 Generation of RNA read counts

For RNASeq analysis, raw fastq file reads were evaluated for artifact sequence using BBDuk (https://sourceforge.net/projects/bbmap/), by kmer matching (kmer=25), allowing 1 mismatch and detected artifact was trimmed from the 3' end of the reads. RNA spike-in reads, PhiX reads and reads containing any Ns were removed. Quality trimming was performed using the phred trimming method set at Q6. Finally, following trimming, reads under the length threshold were removed (minimum length 25 bases or 1/3 of the original read length - whichever is longer). Raw reads from each library were aligned to the reference genome using TopHat (Kim *et al.*, 2013) with only unique mapping allowed (BAMs/ directory). If a read mapped to more than one location, it was ignored. HTSeq (Anders *et al.*, 2014) was used to generate the raw gene counts.

References

Anders S, Pyl PT, Huber W. **2014**. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. **2009**. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Darling A, Carey L, Feng W. **2003**. The Design, Implementation, and Evaluation of mpiBLAST. *Proc Cluster World* **2003**.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. **2000**. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* **300**: 1005–1016.

Fleischmann A, Krings M, Mayr H, Agerer R. **2007**. Structurally preserved polypores from the Neogene of North Africa: *Ganodermites libycus* gen. et sp. nov. (Polyporales, Ganodermataceae). *Review of Palaeobotany and Palynology* **145**: 159–172.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**: 644–652.

Herpoël I, Moukha S, Lesage-Meessen L, Sigoillot J-C, Asther M. **2000**. Selection of *Pycnoporus cinnabarinus* strains for laccase production. *FEMS Microbiology Letters* **183**: 301–306.

Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. **2007**. WoLF PSORT: protein localization predictor. *Nucleic Acids Research* **35**: W585–W587.

Katoh K, Standley DM. **2013**. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. **2013**. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**: R36–R36.

Martin J, Bruno VM, Fang Z, Meng X, Blow M, Zhang T, Sherlock G, Snyder M, Wang Z. **2010**. Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. *BMC Genomics* **11**: 663.

Pellegrin C, Morin E, Martin FM, Veneault-Fourrey C. **2015**. Comparative Analysis of Secretomes from Ectomycorrhizal Fungi with an Emphasis on Small-Secreted Proteins. *Frontiers in Microbiology* **6**: 1278.

Petersen TN, Brunak S, von Heijne G, Nielsen H. **2011**. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**: 785–786.

Puttick MN. **2019**. MCMCtreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. *Bioinformatics* **35**: 5321–5322.

Stamatakis A. **2014**. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Stichting C, Centrum M, Dongen S. **2000**. A Cluster Algorithm for Graphs. *Report-Information systems* **10**.

Varga T, Krizsán K, Földi C, Dima B, Sánchez-García M, Sánchez-Ramírez S, Szöllősi GJ, Szarkándi JG, Papp V, Albert L, *et al.* 2019. Megaphylogeny resolves global patterns of mushroom evolution. *Nature Ecology & Evolution* **3**: 668–678.

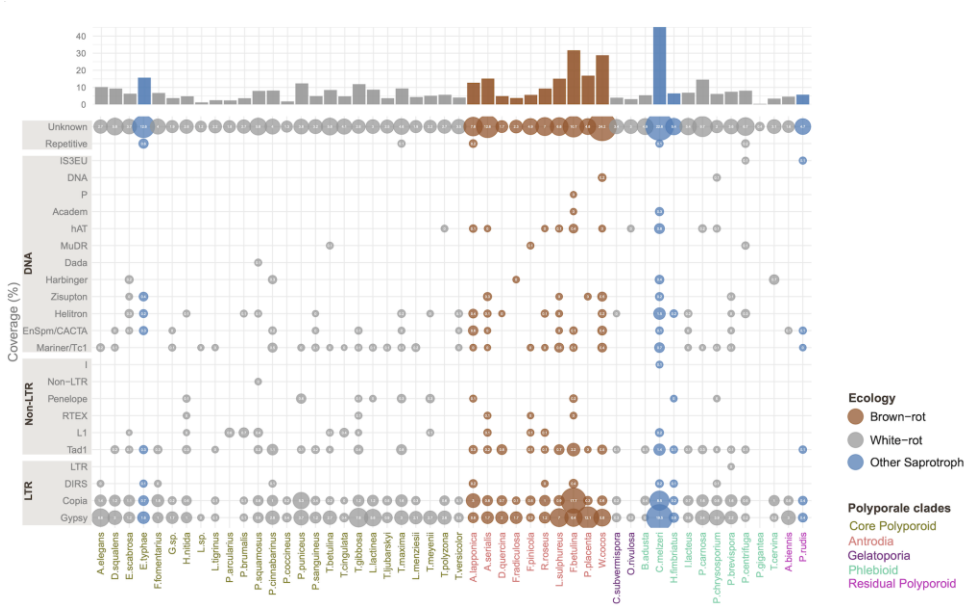


Figure S1: The coverage of transposable elements identified in the 50 genomes.

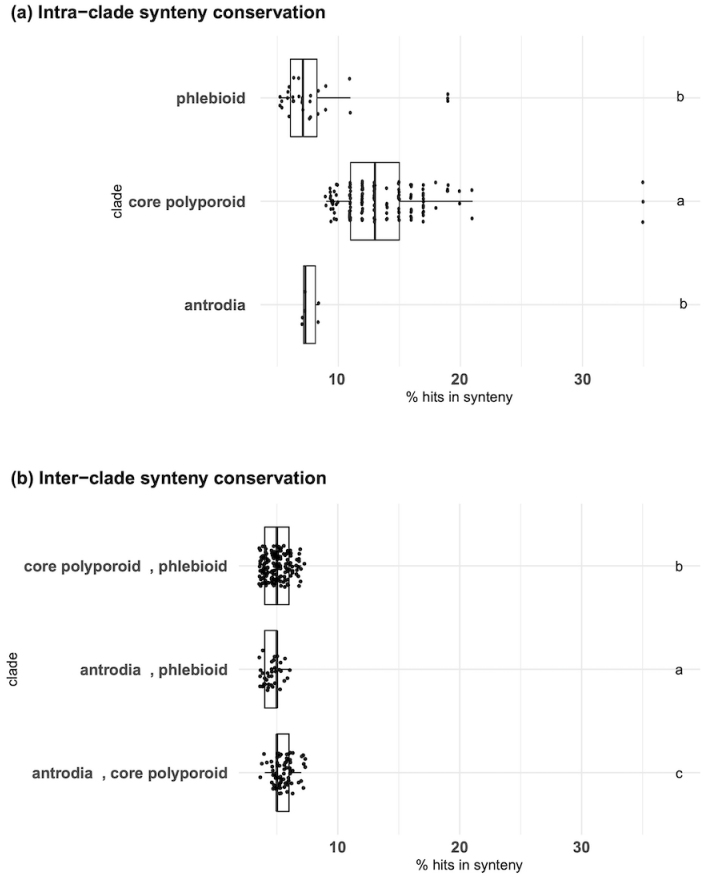


Figure S2: Boxplots showing distribution of synteny conservation within and between clades.

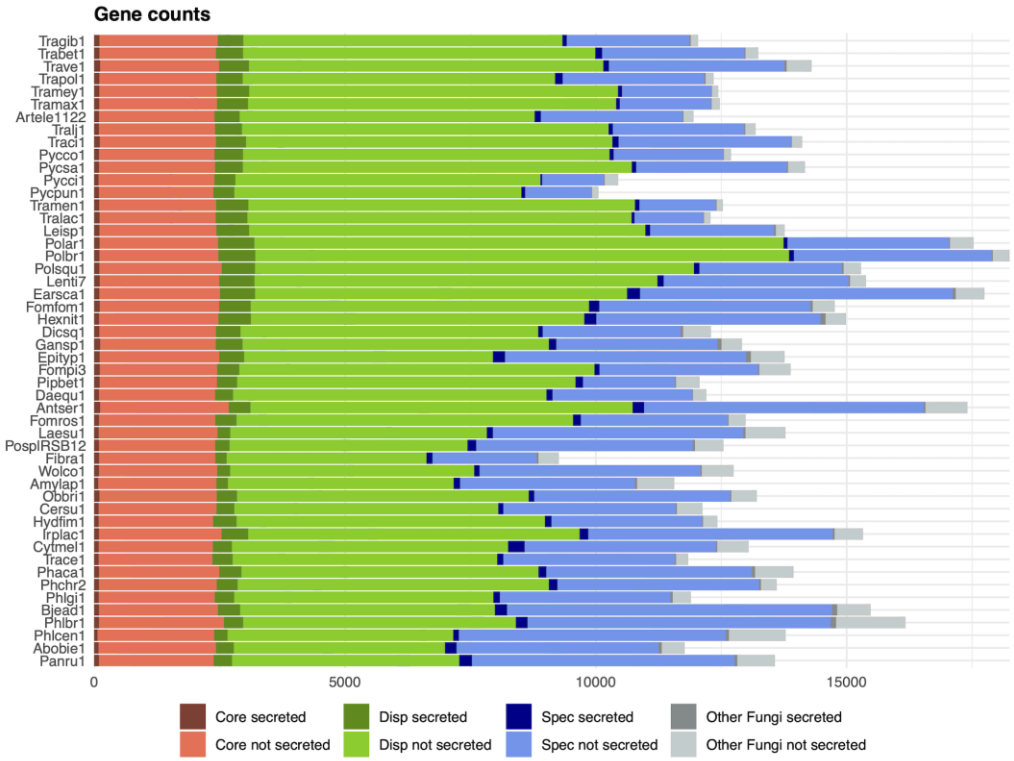


Figure S3 Protein conservation across Polyporales.

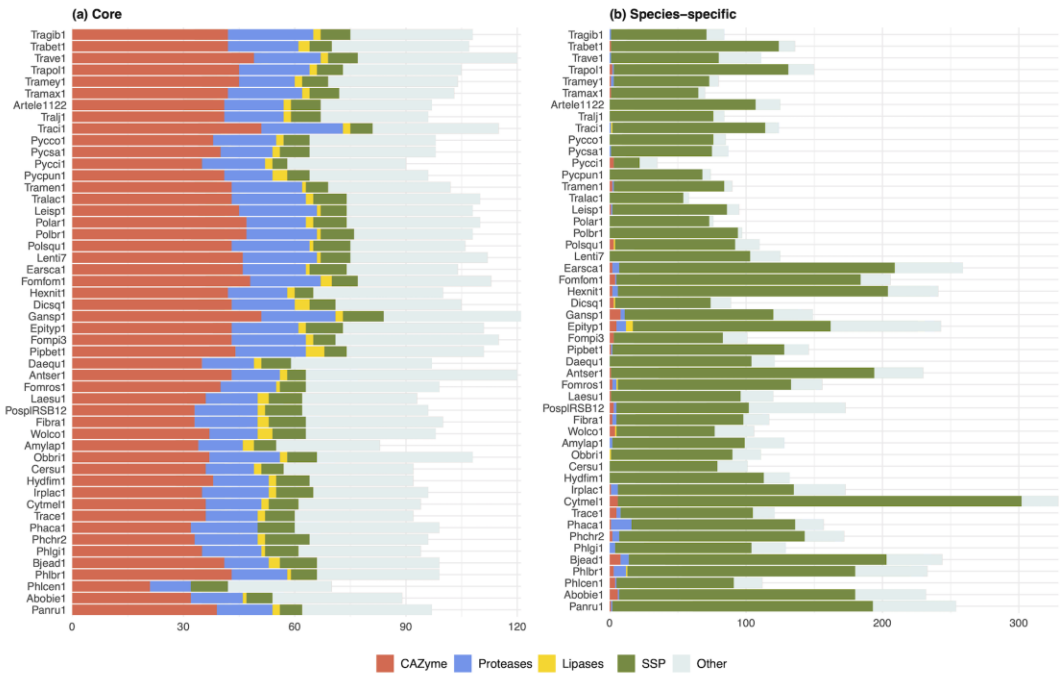


Figure S4 Counts of genes coding for predicted secreted CAZymes, proteases, lipases and SSP among (a) the core secreted proteins and (b) the species-specific secreted proteins in each of the 50 Polyporales genome.

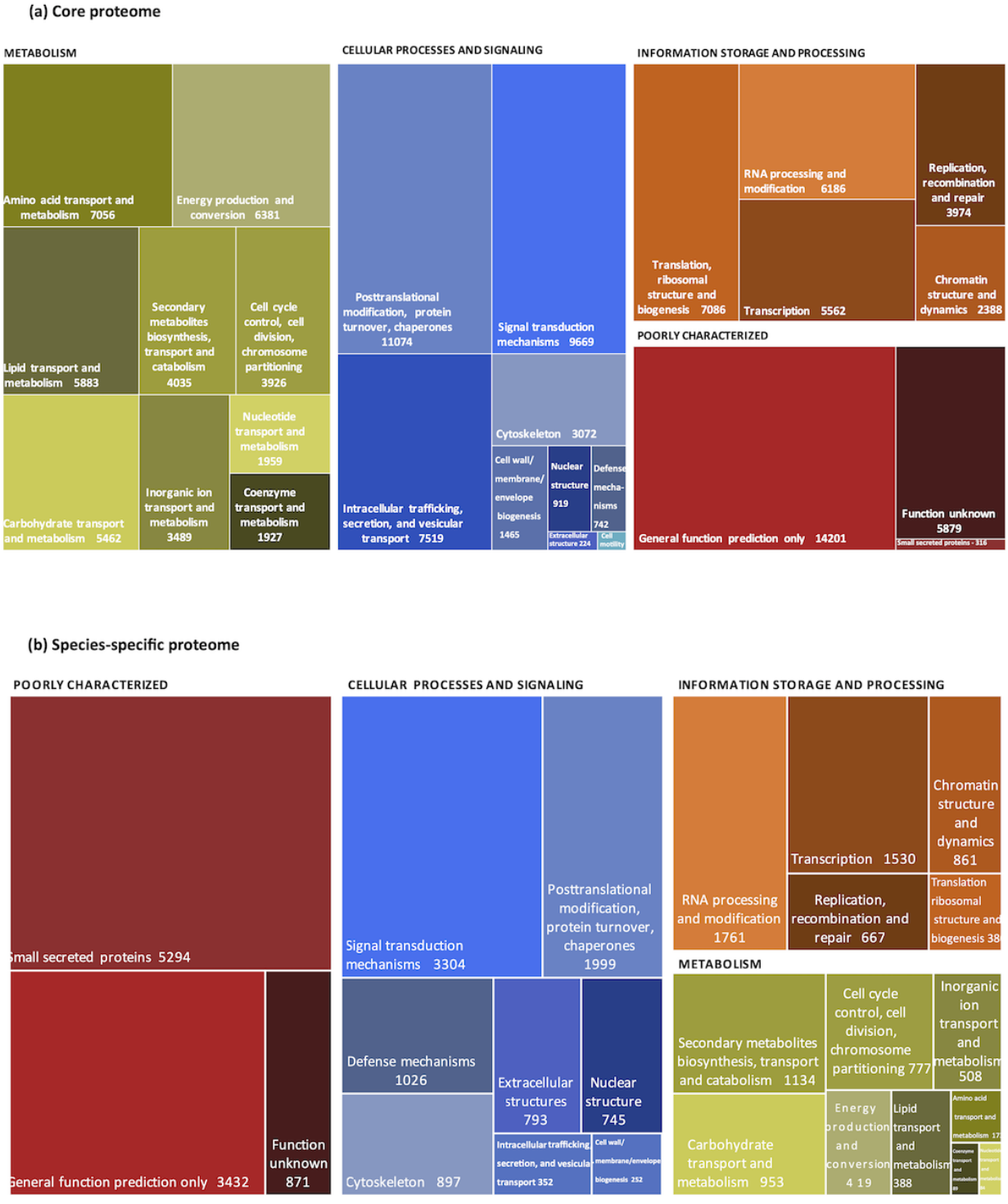


Figure S5 Overall differences of KOGG categories between core (a) and species-specific (b) Polyporales proteomes.

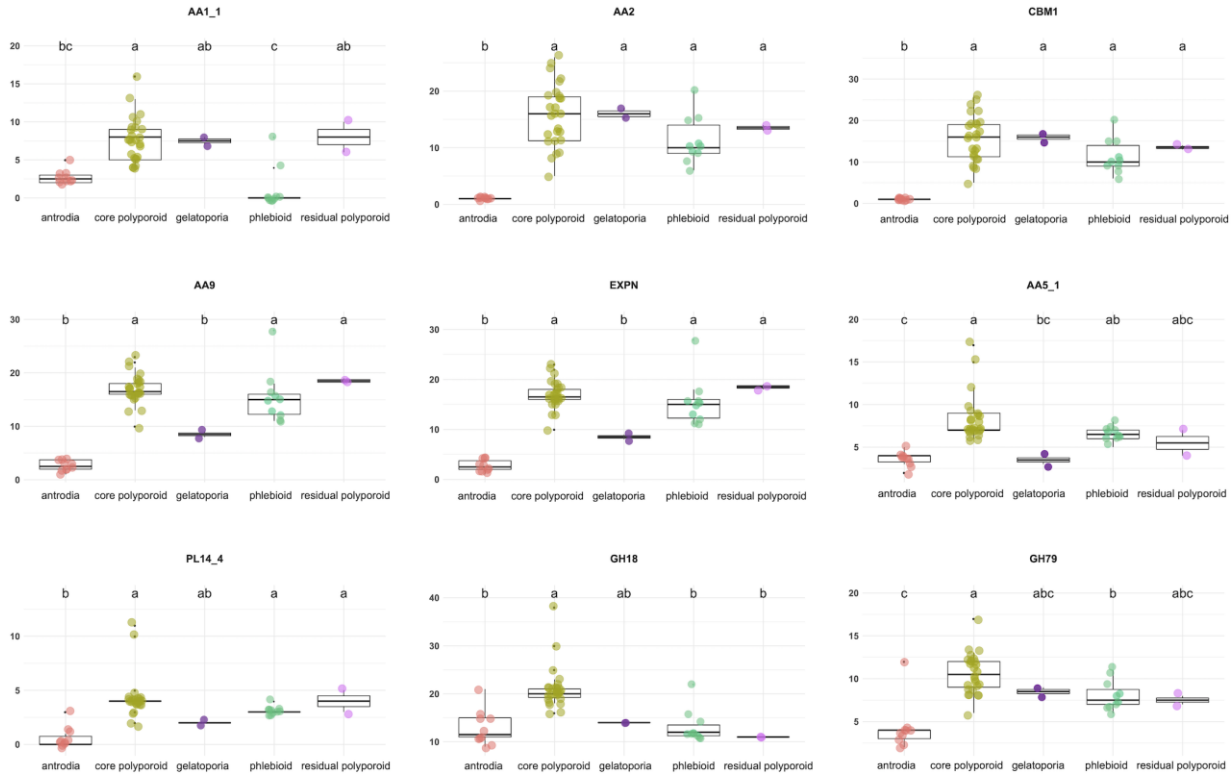


Figure S6 Boxplots showing counts of selected secreted CAZymes in each clade.

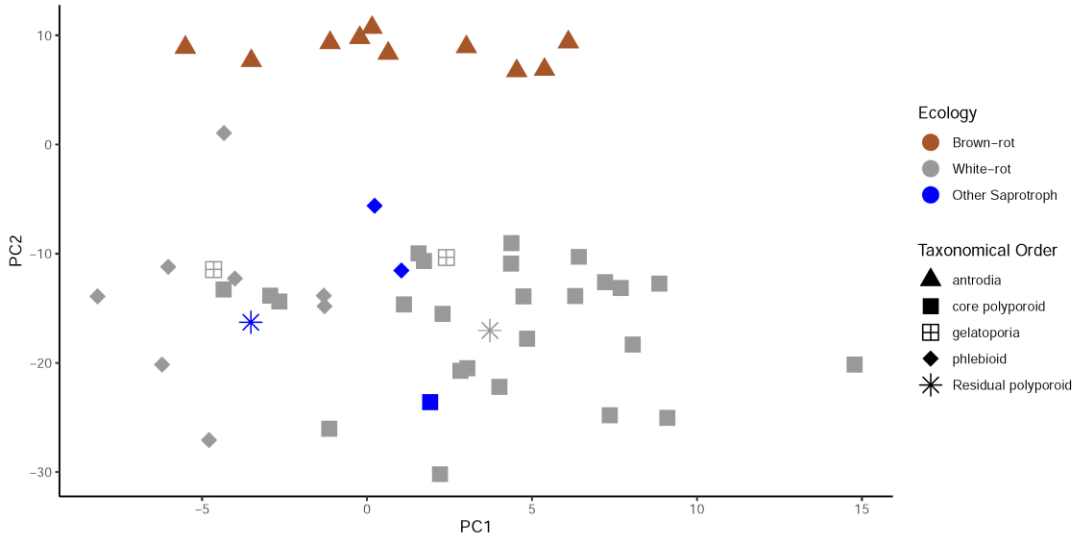


Figure S7 PhyloPCA of the counts of genes coding for secreted CAZymes in 50 Polyporales species using phyl.pca function from phytools in R package.

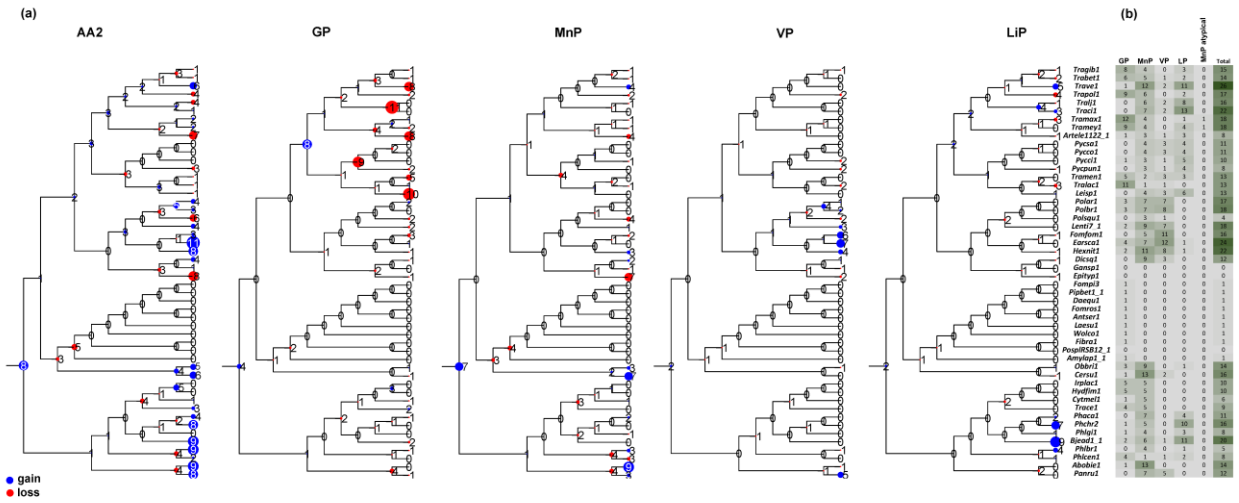


Figure S8 Evolutionary history of gene loss and gains for Class II peroxidases (AA2) from each type; general peroxidase (GP), manganese peroxidase (MnP), versatile peroxidase (VP) and lignin peroxidase (LiP) (a), and their respective gene counts in Polyporales genomes (b).

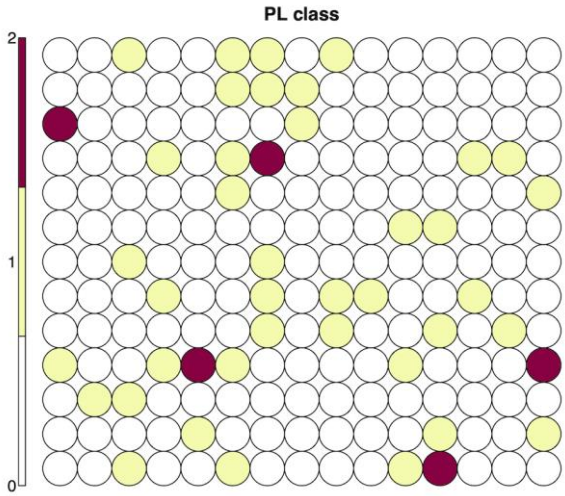


Figure S9 Overview of the transcription regulation of genes coding for polysaccharide lyases from six Polyporales species.

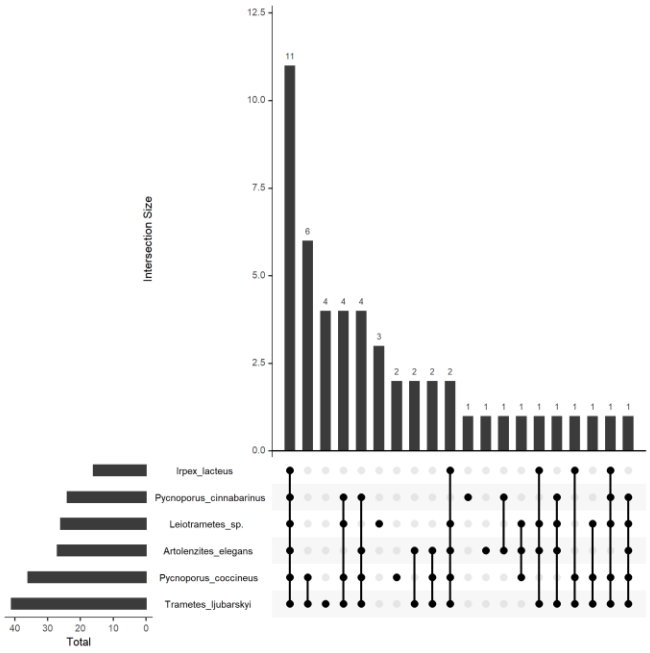


Figure S10 Number of CAZy families with shared up-regulation after 3 day-growth on cellulose, aspen, pine, or wheat straw in comparison to maltose, in the six tested Polyporales species.

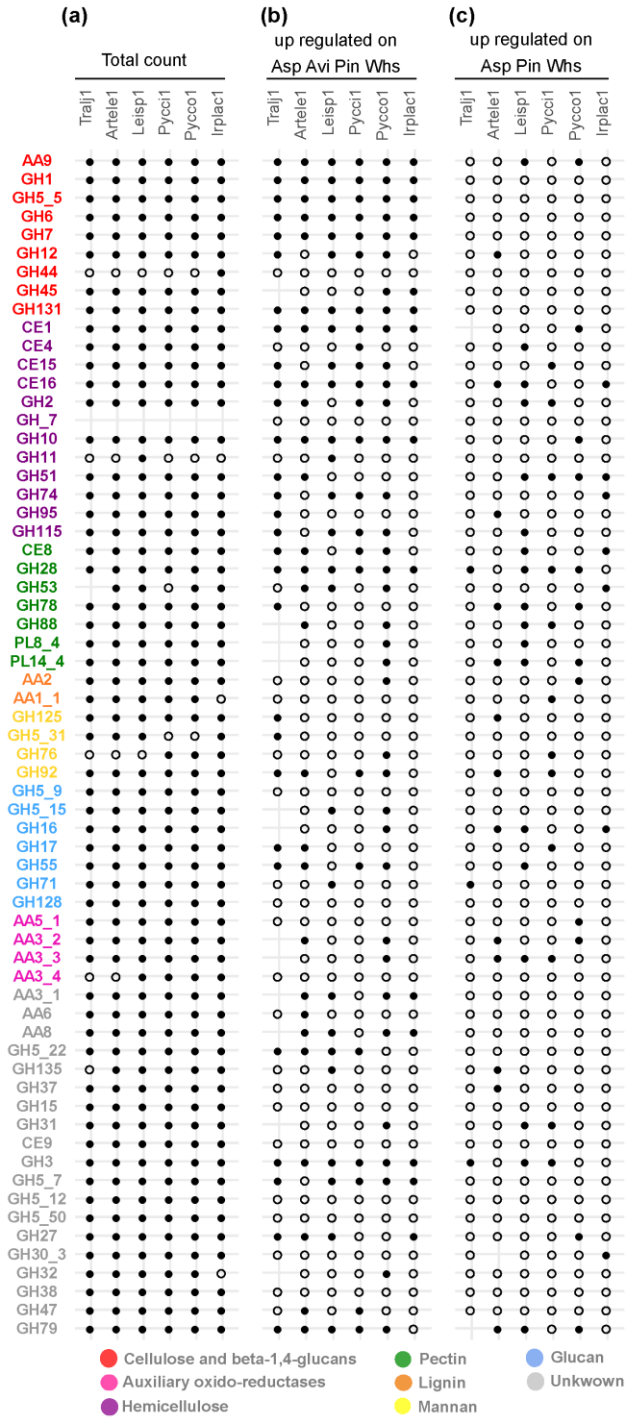


Figure S11 CAZy families for which at least one gene was up-regulated after 3 day-growth on each substrate.

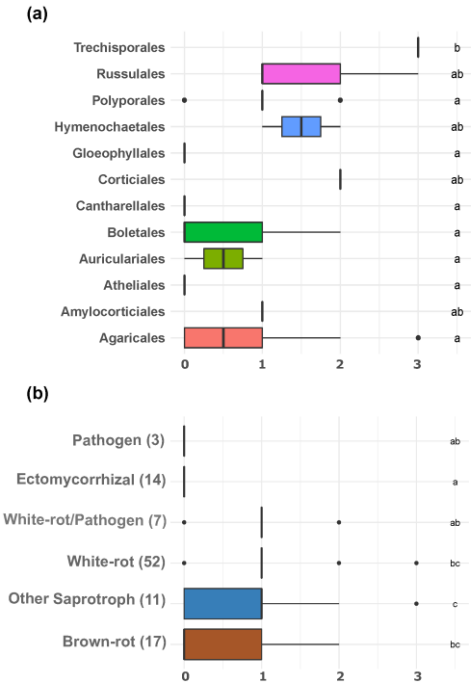


Figure S12 Numbers of AA9 LPMOs genes with X282 extension in the genomes of Agaricomycetes from different orders (a) and lifestyles (b).

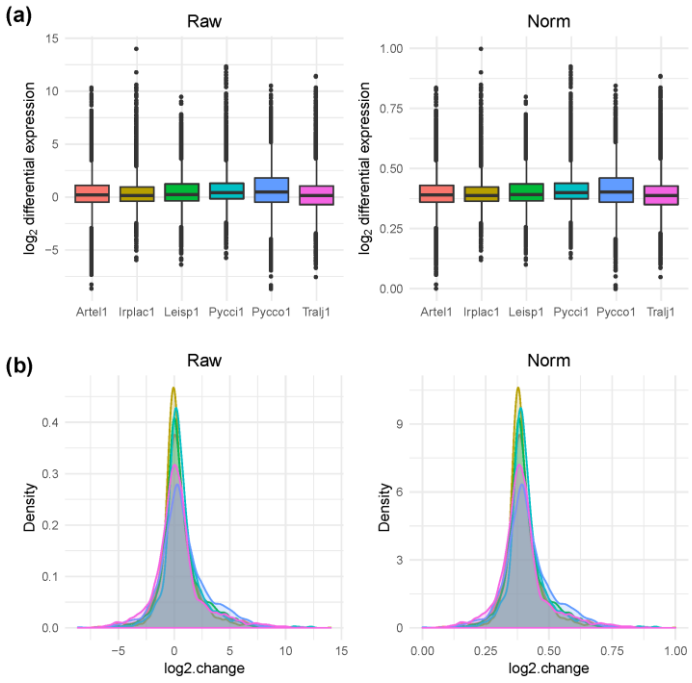


Figure S13 Comparisons of the distributions between the log2 fold change in transcript levels after DESeq2 normalization of transcript read counts (a) and subsequent unity-based normalization (b) for the whole transcriptomes of the 6 species.

IV.3. An HMM approach to genome-wide identification of fungal sesquiterpene synthases.

Chapter IV.3 summary:

Terpenes and terpenoids constitute a large and structurally diverse class of natural products involved in a wide range of biological functions in all living organisms, notably plants, bacteria and fungi. In Fungi, the biosynthesis of terpene backbones starts from acetyl-coA with the mevalonate pathway and ends with the cyclization of a linear hydrocarbon via a cyclase enzyme. Sesquiterpene synthases (STS), the most abundant and diverse group of terpene cyclases in fungi, cyclize the farnesyl diphosphate precursor that has 15 carbon atoms. Previous studies in other groups suggested that fungal STS could be classified into 4 classes based on 4 possible cyclisation mechanisms.

Several bioinformatic tools have been developed to mine genome sequences and identify the genes involved in the sequential steps of terpene biosynthesis. However, those tools were primarily built using plant terpene synthase sequences which show low sequence similarities with the fungal enzymes and are not optimized for mining fungal genomes.

During my visit to Igor Grigoriev lab, the JGI Fungal Genomics group (Berkeley, California), I have developed a pipeline for the accurate identification of fungal STS. Starting with a limited set of available characterized STS from fungi, we expanded the set of sequences for each cyclization type by identifying candidate STS from the order Polyporales. We used these enlarged sets of predicted STS to build for each cyclization type, a first generation HMM model. Those models served to identify candidate STS in the 1,420 available fungal genomes. We then used the candidate STS having the highest HMM scores to build more robust second, and third generation HMM models. The candidate STS from the whole fungal kingdom were identified with the robust HMM models and classified into classes related to their predicted cyclisation mechanisms.

This work allowed the identification of 11,085 candidate STS genes in 1,420 fungal genomes, among which 6,207 were not identified by Pfam search, and the prediction for their cyclisation mechanisms.

This work is presented in the manuscript “An HMM approach to genome-wide identification of fungal sesquiterpene synthases”, by Hayat Hage, Asaf Salamov, Gilles Iacazio, Igor grigoriev, Marie-Noëlle Rosso. (*In preparation*.)

An HMM approach to genome-wide identification of fungal sesquiterpene synthases.

Hayat Hage[1], Asaf Salamov[3], Gilles Iacazio[2], Igor grigoriev[3,4,5], Marie-Noëlle Rosso[1].

1 INRAE, Aix Marseille Univ, UMR1163, Biodiversité et Biotechnologie Fongiques, 13009, Marseille, France

2 Aix-Marseille Univ, CNRS, Centrale Marseille, iSm2, Marseille, France.

3 US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

4 Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

5 Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA, USA

Keywords: terpene cyclase, terpene synthase, basidiomycete fungi, ascomycete fungi, early diverging fungi

Abstract

Terpenes and terpenoids are secondary metabolites that play important roles in basal cellular functions, but also in chemical interactions or in protection against abiotic and biotic environmental stresses. In fungi, these compounds are involved in the communication with the environment. Fungi produce a wide range of sesquiterpenes, which result from the cyclization of the C15 linear precursor, farnesyl diphosphate, by sesquiterpene synthases (STS). Accordingly, STS are more abundant than mono- and di-terpene cyclases in Ascomycota and Basidiomycota. Previous results from the literature suggested that the initial steps of cyclization by basidiomycete STS and the resulting structures of the terpene backbones can be predicted from the enzyme protein sequence. However, much is still to be done for accurate identification of STS in fungal genomes, due to limited conservation of fungal sequences with plant sequences. Here we developed a new set of HMM profiles for the identification of fungal STS and the prediction of the cyclization initiation steps. A first set of HMM profiles was defined based on 1) the protein sequences of characterized fungal STS, 2) a set of candidate STS from 24 Polyporales genomes, 3) the classification of Polyporales STS according to protein sequence similarity and the presence of motives able to discriminate enzymes with four different cyclization mechanisms. This set of HMM profiles was used for mining 1,420 fungal genomes, from which the sequences with the best HMM hits were retrieved to build a set of 2nd generation HMM profiles. By using these HMM profiles, 11,085 candidate STS genes were identified, among which 6,207 (56%) were not identified by the currently available tools and classified according to the predicted preferential cyclization mechanism.

Introduction

Fungi are known to be prolific producers of natural compounds with relevant pharmacological and physiochemical properties (drugs, flavors, pigments, fragrances, …) (Bohlmann & Keeling, 2008; Xiao & Zhong, 2016; Christianson, 2017; Mafu & Zerbe, 2018). Among those compounds, terpenes and terpenoids constitute the largest and most structurally diverse class, with over 80,000 different molecules described (Christianson, 2017). Terpenes are ubiquitous and play essential roles, for example in membrane structure and function (e.g. steroid and sterol precursors) or in respiration chain electron transport (e.g. ubiquinone and menaquinone precursors). Fungi also produce terpenes and terpenoids to interact with bacteria or other fungi living in the same ecological niche (Alexander *et al.*, 2009; O'Leary *et al.*, 2019), to establish associations with plants (Li *et al.*, 2016), to attract insects that enhance spore dissemination (Steinebrunner *et al.*, 2008) or to repel harmful ones (Rohlfs & Churchill, 2011; Sivanandhan *et al.*, 2018). Besides, several fungal terpenes have shown medicinal anticancer activity (Xiao & Zhong, 2016), such as the Δ6-protoilludene derived sesquiterpenoids produced by both basidiomycetes and ascomycetes (de Sena Filho *et al.*, 2016).

In fungi, terpenes are produced via the mevalonate pathway, in which acetyl-CoA is transformed into the two universal precursors, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) in five and six enzymatic steps respectively. The head-to-tail fusion of the C-5 DMAPP with one, two, or three units of C-5 IPP forms the linear precursors of terpenes, geranyl diphosphate (GPP; the C10 precursor of monoterpenoids), farnesyl diphosphate (FPP; the C15 precursor of sesquiterpenoids) or geranylgeranyl diphosphate (GGPP; the C20 precursor of diterpenoids), respectively. The enormous structural diversity of terpenes comes from the cyclization the linear precursors into polycyclic terpene structures by terpene synthases (or terpene cyclases).

In fungi, sesquiterpene synthases (STS), which catalyze the cyclization of FPP, are more abundant than mono- and di-terpene synthases (Schmidt-Dannert C, 2014), consistent with the abundance and diversity of sesquiterpenes detected in the volatile organic compounds produced by fungi in response to biotic stress (Hiscox & Boddy, 2017). To the best of our knowledge, about a hundred fungal STS have been characterized to date. In ascomycetes, efforts have been focused on enzymes involved in the synthesis of mycotoxins, such as DON (deoxynivalenol) (e.g. trichodiene synthase from *Fusarium graminearum* (Proctor *et al.*, 1995)), trichodermin (e.g. trichodiene synthase from Trichoderma brevicompactum (Tijerino *et al.*, 2011)), the PR toxin (aristolochene synthase from *Penicillium roqueforti* (Cheeseman *et al.*, 2014)), or on crop pests (e.g. *Colletotrichum acutatum* (Buchvaldt Amby *et al.*, 2016) and *Botrytis cinerea* (Wang *et al.*, 2009). In addition, a few STS have been characterized from

endophyte fungi, which attract interest for secondary metabolites that have toxicity against insect pests (Wu *et al.*, 2016). In basidiomycetes, recent efforts have been largely dedicated to the search of STS involved in the synthesis of Δ6-protoilludene, a precursor of anticancer molecules, for example in edible fungi (e.g. *Agrocybe aegerita* (Zhang *et al.*, 2020), in tree pathogens (*Armillaria gallica* (Engels *et al.*, 2011)) and *Heterobasidion annosum* (Zhang *et al.*, 2020)) and in wood decay fungi ((e.g. *Omphalotus olearius* (Wawrzyn *et al.*, 2012), *Postia placenta* (Ichinose & Kitaoka, 2018) or *Stereum hirsutum* (Nagamine *et al.*, 2019)), among others.

Yet, much is still to be done for accurate identification of fungal STS and their biochemical characterization. Genome mining for terpene synthases in fungi is limited by the low sequence similarity between plant and microbial enzymes outside the metal binding motifs [D(D/E/N)XX(D/E)] and NSE/DTE (Zhang *et al.*, 2020). The conserved suites of amino acids (Hidden Markov Models; HMM profiles) classified in the Pfam database, primarily built from plant terpene synthases, and the tools based on the identification of biosynthesis gene clusters have shown limited success for fungal genome mining (Khaldi *et al.*, 2010; Weber *et al.*, 2015). Consequently, there is a critical need for the development of new tools to accurately identify fungal terpene synthases genes.

STS catalyze the ionization-dependent removal of inorganic pyrophosphate (PPi) from (2E,6E)-FPP, and generate a reactive carbocation. Further 1,10 or 1,11 cyclization generate a E,E-germacradienyl cation or trans-humulyl cation, respectively. Alternatively, an initial isomerization of (2E,6E)-FPP may form (3R)-nerolidyl pyrophosphate (NPP), further modified via 1,6; 1,7; 1,10 or 1,11cyclization, leading to a bisabolyl, cycloheptanyl, Z,E-germacradienyl or (Z,E)-humulyl cation, respectively (E.M. & R., 2000). The resulting carbocations can undergo further cyclization reactions, hybrid transfers or methyl migrations, increasing the diversity of carbon skeletons that can be generated by STS (Christianson, 2006). Interestingly, recent phylogenetic analyses on characterized STS from basidiomycetes suggested that STS with similar cyclization mechanism cluster together (Flynn & Schmidt-Dannert, 2018; Nagamine *et al.*, 2019; Zhang *et al.*, 2020). From these studies, four STS clades were identified that shared a same cyclization mechanism; clade 1 STS catalyze the 1,10-cyclization of (2E,6E)-FPP carbocation, clade 2 STS catalyze the 1,10-cyclization of the (3R)-NPP carbocation; clade 3 STS catalyze the 1,11 cyclization of (2E,6E)-FPP carbocation (trans-humulyl-type cyclases) and clade 4 STS catalyze the 1,6 or 1,7 cyclization of the (3R)-NPP carbocation. Therefore, based on their sequence similarities, STS genes that catalyze a same cyclisation mechanism could be grouped together into one of the 4 class. With the enormous amount of genome sequence information available to date, the speed of terpene synthase mining exceeds largely that of their

biochemical characterization and predicting a possible final product from the enzyme sequence will remain of great interest until chemical studies catch up with genomics.

In this work, we have developed a powerful method for the accurate identification of STS genes in fungal genomes. Using a set of sequences from characterized basidiomycete STS, we first expanded the set of sequences for each cyclization type by identifying 309 candidate STS from the order Polyporales. We used these enlarged sets of predicted STS to build a first generation HMM model for each cyclization type. We used the STS that had the highest scores with the first generation HMM models to build more robust second generation HMM models. The second generation HMM models allowed the identification of 11,085 candidate STS genes in 1,420 fungal genomes, among which 6,207 were not identified by Pfam search, and proposed a prediction for their cyclization mechanism. Our results suggest that the Basidiomycota phylum encodes 2-3 times larger STS families than Ascomycota do, possibly caused by way higher tandem duplication events in Basidiomycota.

Materials and Methods

Identification of STS genes in Polyporales

We selected a subset of characterized STS genes from each clade defined in Schmidt-Dannert et al. (2014) as queries against Polyporales genomes (Table S1). Thus, we chose as representatives of (i) clade 1: Cop2 and Omp3 from the Agaricales *Coprinopsis cinerea* and *Omphalotus olearius* respectively, (ii) clade 2: Stehi1_ 128017 from the Russulales *Stereum hirsutum* and Omp5a*,* (iii) Clade 3*:* BcBOT2 from the ascomycete *Botrytis cinerea* and Omp6, and finally (iv) clade 4: Ffsc6 from the ascomycete *Fusarium fujikuroi*, Omp10 and Fompi84944 from the Polyporales *Fomitopsis pinicola*. The sequences were retrieved from GenBank database and queried using blastp from ncbi-blast+, version 2.2, against 24 Polyporales genomes (Table S2; Hage et al., in prep.). The STS hits that matched the following criteria were retrieved: 1) %identity > 30%, %query coverage > 60%, 2) conservation of the metal-binding domains [(D/N)DXX(D/E) or DDXXXE] in the N-terminal and (NxxxSxxxE) in the C-terminal, 3) sequence length between 250 and 500 amino acids and 4) presence of the terpenoid synthase Superfamily domain (SCOP 48576), which includes the following five families: isoprenyl diphosphate synthase (SCOP 48577), squalene synthase (SCOP 48580), terpenoid cyclase C-terminal domain (SCOP 48583), aristolochene/pentalenene synthase (SCOP 48586) and trichodiene synthase (SCOP 69113). To ensure that no hits were missing, and no false positive were included, we tested the method on

the genome of *Omphalotus olearius*, for which STS have been characterized (Wawrzyn *et al.*, 2012), and compared the list of characterized STS with the list of hits found using our method.

Classification of Polyporales STS genes into 4 clades

To classify the predicted STS genes into one of the 4 clades, we aligned the STS hits found in each genome with the selected characterized STS using MAFFT and computed 24 distance matrix using Fprotdist from the PHYLIP package in EMBOSS. To assign a hit to a clade, 24 dendrograms were built from those matrices using "dendextend" package in R, by using the hits retrieved from each genome and the characterized STS sequences used as queries. We obtained four sets of sequences, for clade 1, 2, 3 and 4, in addition to a fifth set that we called "unclassified", which contained the hits that did not cluster with any of the characterized STS on the dendrograms. We searched signature motifs of each clade using the MERCI tool (Vens *et al.*, 2011). This tool identifies conserved/highly represented motifs in a batch of sequences (here sequences from a specific clade, considered as "positive" sequences), that are absent in the other batches (here the other three clades, considered as "negative" sequences). The phylogenetic analyses of STS sequences were done using Randomized Axelerated Maximum Likelihood (RAxML) with PROTGAMMAAUTO model and 500 bootstrap values.

Generating robust HMM-models for fungal STS identification and classification

The STS sequences from a same clade were aligned using Mafft version v7.429, with option GOP 2.3 and GEP 0.63 (Long *et al.*, 2016). The poorly aligned regions were removed using trimAl 1.2 and the quality of each alignment was manually inspected. Each of the five trimmed alignments was used to build one HMM Hidden Markov Model using the "hmmbuild" command from HMMER3.1. The HMM profiles attributed to each clade were called cladeX-HMM, where "X" is the name of the clade. Those five cladeX-HMM models served to search for STS hits in all fungal genomes available in MycoCosm (1,420 genomes on April 1st 2019), using the hmmsearch command with gathering thresholds = 25. A hit was attributed to a clade according to the cladeX-HMM model that found the hit. If a hit was commonly found by more than one cladeX-HMM, we assigned it to the clade of the model with the highest HMM score. The newly generated sets of sequences gathered all fungal STS sequences in each clade. From each set, the 100 sequences with the highest HMM scores were aligned together as described previously and the trimmed alignments were used to build second generation cladeX-HMM models.

Results and discussion

Comparison of plant vs fungal STS genes

Plant STS sequences contain a 252 aa N-terminal region of unknown biochemical function (Whittington *et al.*, 2002) that is highly conserved, facilitating thus their detection by application of a simple BLAST against plant protein databases. By contrast, fungal terpene synthases lack this conserved N-terminal region and the C-terminal region shows low sequence identity (<28 %) with the plant sequences (Table S3). Consequently, the Pfam domains built from the sequences of plant terpene synthases, e.g. PF01397 and PF03936 for the N-terminal and C-terminal regions respectively, are not suitable for the identification of fungal STS.

Identification of 309 STS genes in Polyporales

As a first step towards the construction of HMM profiles for fungal STS identification, we searched for candidate STS genes in the genomes of 24 Polyporales species (Figure 1).

We first selected nine characterized STS from basidiomycetes, for which the cyclization mechanism on FPP was demonstrated; Cop2 and Omp3 for the 1,10-cyclization of (2E,6E)-FPP (Clade 1 type), Omp5a and Stehi1_ 128017 for the 1,10-cyclization of the (3R)-NPP (clade 2 type), BcBOT2 and Omp6 for the 1,11 cyclization of (2E,6E)-FPP (clade 3 type) and finally Ffsc6, Fompi84944 and Omp10 for the 1,6 or 1,7 cyclization of the (3R)-NPP (clade 4 type). Using a blast search against 24 newly sequenced fungal genomes (Hage et al., in prep.), identified a total of 3,126 candidate STS genes (Table 1). Among those, 454 genes had more than 30% identity and 60% query coverage with the sequences of the characterized enzymes. From those, 318 genes had a SUPERFAMILY domain related to terpene synthases as classified in SCOP 48576, and a total of 309 genes also had the conserved motives for binding the metal ion and a sequence length comprised between 250 and 500 amino acids. Those 309 genes were retrieved as candidate STS genes. To ensure that those filtering steps did not produce false positives and that no hits were missing, we tested the method on the genome of *Omphalotus olearius,* for which we retrieved the nine STS genes previously identified by Wawrzyn *et al.*, 2012.

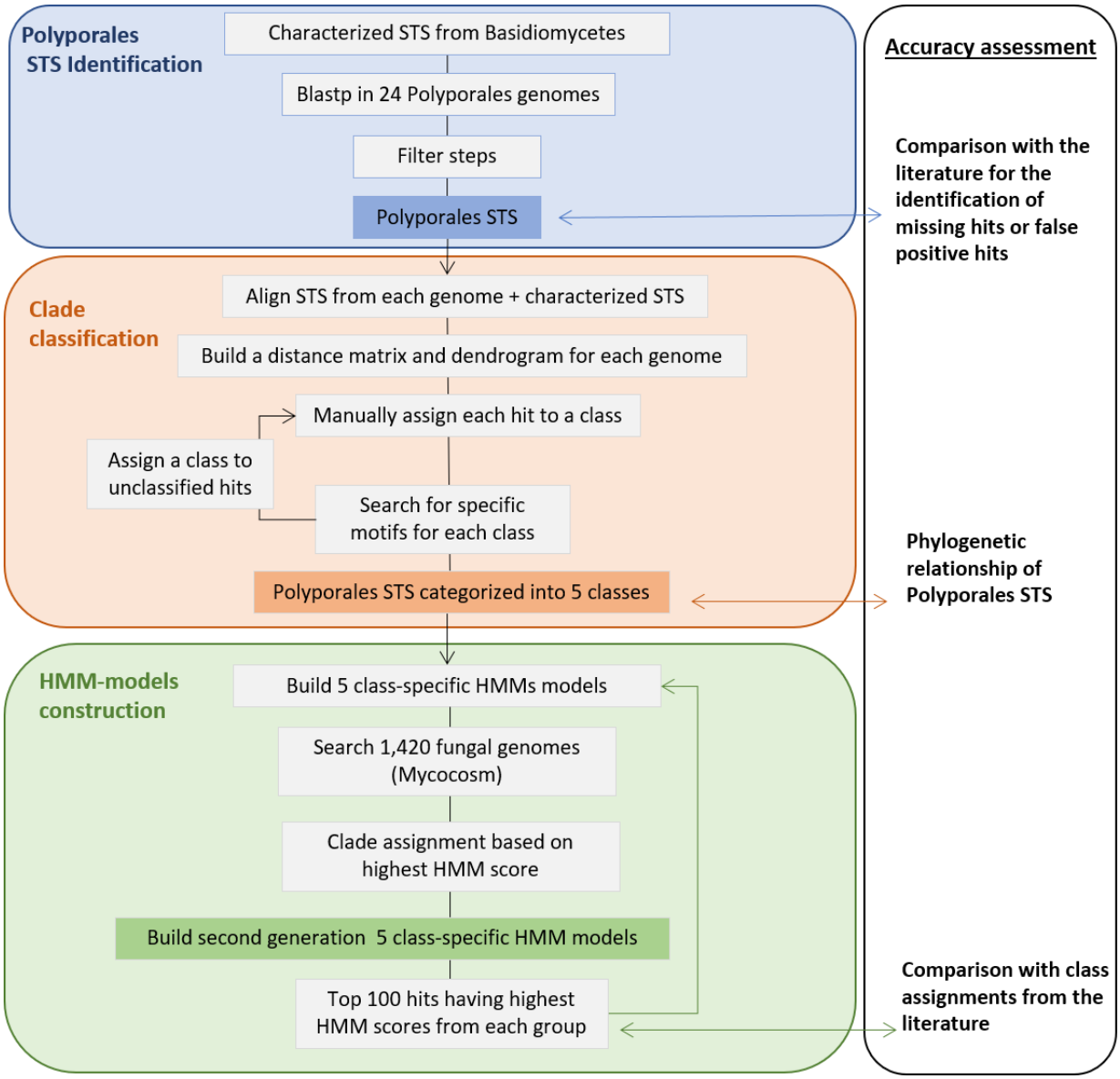


Figure 1 Different steps followed for the construction of final robust HMM-models for fungal STS identification.

Table 1 Count of candidate STS identified in each of the 24 Polyporales genomes, obtained after three filter steps.

Species	Genome	Blast hits	Filter 1 sequence identity > 30% and query coverage > 60%	Filter 2 SUPERFAMILY terpene synthase domain	Filter 3 250 < length (aa) < 500
Abortiporus biennis	Abobie1	133	17	11	10
Artolenzites elegans (1663)	Artel1	125	19	11	11
Earliella scabrosa	Earsca1	178	23	21	20
Fomes fomentarius	Fomfom1	147	20	16	16
Rhodofomes roseus	Fomros1	122	23	17	17
Hexagonia nitida	Hexnit1	153	22	16	16
Irpex lacteus	Irplac1	96	13	6	4
Leiotrametes sp.	Leisp1	176	28	20	20
Fomitopsis betulina	Pipbet1	118	26	19	17
Polyporus brumalis	Polbr1	117	14	11	11
Pycnoporus cinnabarinus (137)	Pycci1	116	10	8	8
Pycnoporus coccineus (310)	Pycco1	110	16	9	9
Pycnoporus puniceus	Pycpun1	119	18	8	8
Pycnoporus sanguineus	Pycsa1	135	20	13	13
Trametes betulina	Trabet1	124	20	13	13
Trametopsis cervina	Trace1	80	15	11	11
Trametes cingulata	Traci1	118	17	11	10
Trametes gibbosa	Tragib1	126	19	12	12
Trametes lactinea	Tralac1	157	18	15	14
Trametes ljubarskyi	Tralj1	123	17	11	10
Trametes maxima	Tramax1	122	17	12	12
Trametes menziesii	Tramen1	167	24	18	18
Trametes meyenii	Tramey1	131	18	14	14
Trametes polyzona	Trapol1	133	20	15	15
	Total	**3126**	**454**	**318**	**309**

Classification of Polyporales STS into 4 clades

The 309 candidate STS we identified in Polyporales genomes were classified according to the predicted cyclization mechanism using sequence similarity with characterized enzymes. For this purpose, we built distance matrices of the candidate STS of each of the 24 genomes with the characterized STS (Fig. 2). The genes that showed no similarity with any characterized STS were grouped into the “unclassified” category. This resulted in 87 Polyporales STS grouped in class 1, 45 in class 2, 48 in class 3, 41 in clade class 4 and 88 unclassified. We then searched for signature motifs specific of each clade. For this purpose, we used the MERCI tool that allows the identification of motifs highly conserved in a set of sequences and totally absent from others (Vens et al., 2011). The motif “DTSGC” was present in 95% of class 1 sequences, “LRRENS” in 100% of class 2 sequences, “DLMN” in 77% of class 3 sequences and finally “FYK” in 92% of class

4 sequences (Fig. 3). Among the sequences primarily “unclassified”, 24 showed the conservation of one of those motifs and were reassigned accordingly to one of the four classes, which resulted in 89 Polyporales sequences classified in class 1, 51 sequences classified in class 2, 61 in class 3, 44 in class 4 and 64 still unclassified sequences. A phylogenetic analysis of all the classified Polyporales STS confirmed that the attribution of the primarily unclassified sequences to a class based on the conservation of the corresponding motifs was correct (Fig. 4). Further studies are now required to elucidate whether those highly conserved motifs are responsible for the expected mechanism of cyclization for each clade. Besides, this phylogenetic tree showed that sequences from clade 4 might have undergone an accelerated rate of molecular evolution compared to enzymes from the other clades, offering thus a possible higher diversity of terpenes products.

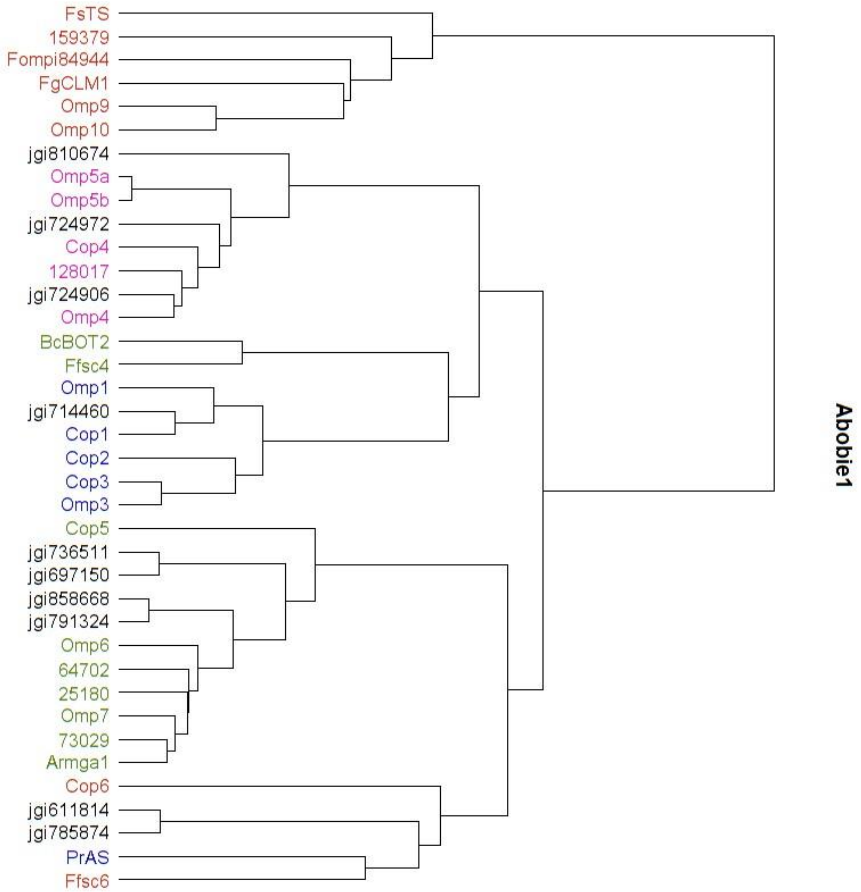


Figure 2 Example of a dendrogram built from distance matrix between identified STS in *Abortiporus biennis* (Abobie1) genome (in black) and STS characterized colored class-wise (class 1 (blue), class 2 (pink), class 3 (green) and class 4 (red)).

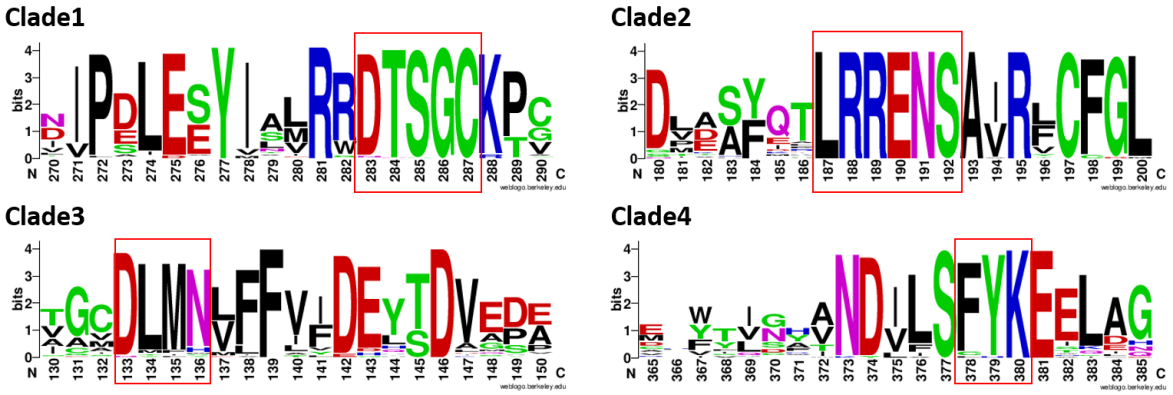


Figure 3 WebLogo of conserved motifs frequently found in STS sequences of each class that are totally absent in sequences from the other classes.

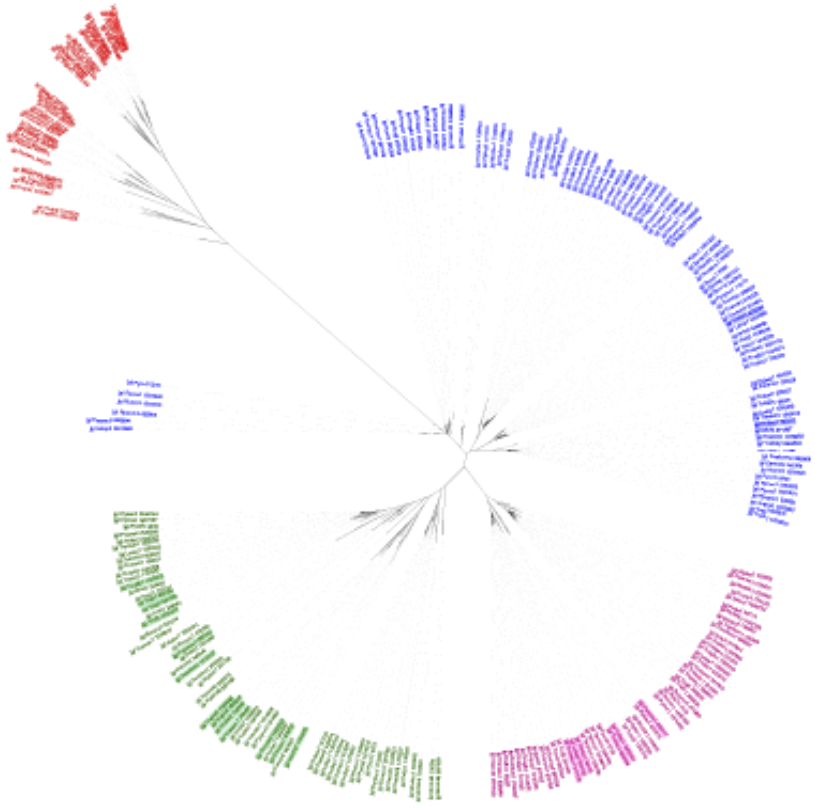


Figure 4 RAxML phylogenetic tree of Polyporales STS genes from class 1 (blue), class 2 (pink), class 3 (green) and class 4 (red). The previously unclassified STS that were reassigned to a class according to the presence of the corresponding conserved motif are highlighted with a clade-wise colored background. The orange branches have bootstrap values > 70.

Construction of HMMs models for each cyclization mechanism

For each class, we built an HMM model using the set of Polyporales sequences assigned to that class, in addition to an HMM model for all the unclassified Polyporales sequences. A search for these five HMM models on 1,420 fungal genomes available on MycoCosm retrieved a total of 11,220 candidate STS genes. The genes retrieved by only one HMM model were assigned to the corresponding class whereas the genes retrieved by the HMM-models of different classes were assigned to the class of the HMM model giving the highest score. This resulted in 2,755 (24%) candidate STS classified in class 1, 2,096 (18%) classified in class 2, 889 (7%) classified in class 3, 2,496 (22%) classified in class 4 and 2,984 (26%) unclassified (Table 2).

Table 2 Counts of STS hits predicted with first and second generation HMM models.

Classification	First generation HMM models	Second generation HMM models
Class 1	2,755	1,996
Class 2	2,096	1,358
Class 3	889	2,488
Class 4	2,496	2,672
Class "unknown"	2,984	2,571
Pfam terpene synthase	4,493	4,488
Other Pfam	614	583
No Pfam	6,312	6,207
Ascomycota (806 genomes)	4,205	4,062
Basidiomycota (466 genomes)	6,980	6,992
Mucoromycota (78 genomes)	9	7
Zoopagomycota (22 genomes)	3	3
Total	11,220	11,085

We then checked if the characterized STS previously analyzed by C. Schmidt-Dannert et al. (2014) were detected by the HMM models and if they were assigned to the right class. All characterized STS previously grouped in clade 1 and clade 2 were detected and assigned to the right class. Among the eight characterized STS from clade 3, four were wrongly assigned to class 1 or class 2. Finally, all the characterized STS from clade 4 were detected and assigned to the right class, except Ffsc6 that was assigned to the “unclassified” class (Table S4).

To improve the assignment of candidate STS to classes, we used the 100 sequences having the highest HMM scores for each profile to build second generation HMM models. A total of 11,085 candidate STS genes were identified in the 1,420 fungal genomes using the second generation HMM models. The clade classification of those 11,085 genes resulted in 1,996 (18%) genes classified in class 1, 1,358 (12%) classified in class 2, 2,488 (22%) classified in class 3, 2,672 (24%) classified in class 4 and 2,571 (23%) unclassified (Table 2). The characterized STS that were assigned to wrong classes with the previous HMM models were reassigned to the correct class with the new HMM models, except for Ffsc4 which was still assigned to class1 (Table S4). A third generation of HMM models didn't allow Ffsc4 to fit the expected classification, and thus the second generation HMM models were used for further analysis.

Among the 11,085 candidate STS genes found by second generation HMM profiles, 40% (4,488 proteins) had a Pfam domain related to terpene synthase and 5% (583 proteins) had Pfam domains related to polyprenyl synthase. Yet 55% (6,207 genes) did not have any Pfam annotation (Table 2). This latter result confirmed that Pfam domains primarily constructed with plant sequences are not suitable for accurate terpene synthase detection in fungal genomes. The presence of a prenyltransferase Pfam domain in some of the newly identified STS can be explained by prenyltransferases featuring a DDxxD motif for binding divalent metal ions and sharing some sequence similarity with class I terpene synthases (Christianson et al., 2017).

Tandem duplications have contributed to increased STS portfolios in Basidiomycota

In total, 63% of the candidate STS were retrieved from Basidiomycetes genomes, 36% from ascomycetes and only 0.1% (10 genes) from early diverging fungi. An average of 18 STS genes/genome was found in basidiomycetes compared to six STS genes/genomes in ascomycetes.

Considering the fact that the initial HMM models were built from Polyporales sequences, we wondered if the HMM profiles could be biased and better match basidiomycete sequences. We constructed five ascomycete HMM models using the top 100 ascomycete sequences having the highest HMM scores from each clade. A search for candidate STS in 1,420 genomes using these Ascomycota-derived HMMs did not allow the identification of additional candidate STS. We thereby validated that, regardless of the taxonomical origin of the STS used to build the HMM models, basidiomycetes encode in general more STS genes compared to ascomycetes. These findings were in line with previous studies that suggested Basidiomycota rely mostly on terpenoids as their predominant natural products, while Ascomycota have evolved a different

arsenal of natural products, with the predominant production of polyketides (PK) and non-ribosomal peptides (NRP) (Boettger & Hertweck, 2013).

To understand this discrepancy between basidiomycete and ascomycete STS gene counts, we analyzed the occurrence of STS gene duplications. We considered genes were in tandem when a second copy was located among the four genes upstream or downstream a first copy on the same scaffold. We observed that 25% of the STS genes in basidiomycetes were organized in tandem repeats compared to only 3% for ascomycetes STS genes. This result suggested that tandem duplications have played major roles in the expansion of STS gene families in basidiomycetes. Yet, following duplication, each gene may evolve in several ways, retaining the same function as its ancestral copy, gaining a new function or becoming a nonfunctional gene (Ohno 1970). Here we observed that the 89% of the STS genes organized in tandem repeat were grouped in the same class, and consequently were predicted to initiate the same cyclization mechanism, and 11% were grouped in different classes and predicted to initiate different cyclization mechanisms. However, the classification of STS and the prediction for the first cyclization steps is not sufficient to predict the final product of STS. STS with slight protein sequence modifications might generate different final terpene products (Garms *et al.*, 2012) and further studies are necessary to assess the diversity of terpenes produced by tandem genes. In addition, gene duplications might lead to the diversification of gene expression profiles, which can now be investigated.

References

Alexander NJ, Proctor RH, McCormick SP. **2009**. Genes, gene clusters, and biosynthesis of trichothecenes and fumonisins in *Fusarium*. *Toxin Reviews* **28**: 198–215.

Boettger D, Hertweck C. **2013**. Molecular diversity sculpted by fungal PKS-NRPS hybrids. *Chembiochem : a European journal of chemical biology* **14**: 28–42.

Bohlmann J, Keeling CI. **2008**. Terpenoid biomaterials. *The Plant journal : for cell and molecular biology* **54**: 656–669.

Buchvaldt Amby D, Manczak T, Petersen MA, Sundelin T, Weitzel C, Grajewski M, Simonsen HT, Jensen B. **2016**. Role of the *Colletotrichum acutatum* sesquiterpene synthase CaTPS in the biosynthesis of sesquiterpenoids. *Microbiology (Reading, England)* **162**: 1773–1783.

Cheeseman K, Ropars J, Renault P, Dupont J, Gouzy J, Branca A, Abraham A-L, Ceppi M, Conseiller E, Debuchy R, *et al.* 2014. Multiple recent horizontal transfers of a large genomic region in cheese making fungi. *Nature communications* **5**: 2876.

Christianson DW. **2006**. Structural Biology and Chemistry of the Terpenoid Cyclases. *Chemical Reviews* **106**: 3412–3442.

Christianson DW. **2017**. Structural and Chemical Biology of Terpenoid Cyclases. *Chemical Reviews* **117**: 11570–11648.

E.M. D, R. C. **2000**. Cyclization Enzymes in the Biosynthesis of Monoterpenes, Sesquiterpenes, and Diterpenes. *Biosynthesis*.

Engels B, Heinig U, Grothe T, Stadler M, Jennewein S. **2011**. Cloning and characterization of an Armillaria gallica cDNA encoding protoilludene synthase, which catalyzes the first committed step in the synthesis of antimicrobial melleolides. *The Journal of biological chemistry* **286**: 6871–6878.

Flynn CM, Schmidt-Dannert C. **2018**. Sesquiterpene Synthase–3-Hydroxy-3-Methylglutaryl Coenzyme A Synthase Fusion Protein Responsible for Hirsutene Biosynthesis in *Stereum hirsutum. Applied and Environmental Microbiology* **84**: e00036-18.

Garms S, Chen F, Boland W, Gershenzon J, Köllner TG. **2012**. A single amino acid determines the site of deprotonation in the active center of sesquiterpene synthases SbTPS1 and SbTPS2 from Sorghum bicolor. *Phytochemistry* **75**: 6–13.

Hiscox J, Boddy L. **2017**. Armed and dangerous – Chemical warfare in wood decay communities. *Fungal Biology Reviews* **31**: 169–184.

Ichinose H, Kitaoka T. **2018**. Insight into metabolic diversity of the brown-rot basidiomycete *Postia placenta* responsible for sesquiterpene biosynthesis: semi-comprehensive screening of cytochrome P450 monooxygenase involved in protoilludene metabolism. *Microbial biotechnology* **11**: 952–965.

Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND. **2010**. SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genetics and Biology* **47**: 736–741.

Li N, Alfiky A, Vaughan MM, Kang S. **2016**. Stop and smell the fungi: Fungal volatile metabolites are overlooked signals involved in fungal interaction with plants. *Fungal Biology Reviews* **30**: 134–144.

Long MC, Deutsch C, Ito T. **2016**. Finding forced trends in oceanic oxygen. *Global Biogeochemical Cycles* **30**: 381–397.

Mafu S, Zerbe P. **2018**. Plant diterpenoid metabolism for manufacturing the biopharmaceuticals of

tomorrow: prospects and challenges. *Phytochemistry Reviews* **17**: 113–130.

Nagamine S, Liu C, Nishishita J, Kozaki T, Sogahata K, Sato Y, Minami A, Ozaki T, Schmidt-Dannert C, Maruyama J-I, *et al.* 2019. Ascomycete *Aspergillus oryzae* Is an Efficient Expression Host for Production of Basidiomycete Terpenes by Using Genomic DNA Sequences. *Applied and environmental microbiology* **85**.

O'Leary J, Hiscox J, Eastwood DC, Savoury M, Langley A, McDowell SW, Rogers HJ, Boddy L, Müller CT. **2019**. The whiff of decay: Linking volatile production and extracellular enzymes to outcomes of fungal interactions at different temperatures. *Fungal Ecology* **39**: 336–348.

Proctor RH, Hohn TM, McCormick SP. **1995**. Reduced virulence of *Gibberella zeae* caused by disruption of a trichothecene toxin biosynthetic gene. *Molecular plant-microbe interactions: MPMI* **8**: 593–601.

Rohlfs M, Churchill ACL. **2011**. Fungal secondary metabolites as modulators of interactions with insects and other arthropods. *Fungal genetics and biology : FG & B* **48**: 23–34.

Schmidt-Dannert C. **2014**. Biosynthesis of Terpenoid Natural Products in Fungi. In: Schrader J., Bohlmann J. (eds) Biotechnology of Isoprenoids. *Advances in Biochemical Engineering/Biotechnology,* **148**.

de Sena Filho JG, Quin MB, Spakowicz DJ, Shaw JJ, Kucera K, Dunican B, Strobel SA, Schmidt-Dannert C. **2016**. Genome of *Diaporthe sp.* provides insights into the potential inter-phylum transfer of a fungal sesquiterpenoid biosynthetic pathway. *Fungal Biology* **120**: 1050–1063.

Sivanandhan S, Ganesan P, Paulraj MG, Ignacimuthu S. **2018**. Larvicidal, Ovicidal, and Histopathological Effects of the Sulphur Polypore Mushroom, *Laetiporus sulphureus* (Agaricomycetes), Collected from Tamil Nadu, India against Mosquitoes. *International journal of medicinal mushrooms* **20**: 1197–1207.

Steinebrunner F, Twele R, Francke W, Leuchtmann A, Schiestl FP. **2008**. Role of odour compounds in the attraction of gamete vectors in endophytic *Epichloë* fungi. *New Phytologist* **178**: 401–411.

Tijerino A, Cardoza RE, Moraga J, Malmierca MG, Vicente F, Aleu J, Collado IG, Gutiérrez S, Monte E, Hermosa R. **2011**. Overexpression of the trichodiene synthase gene tri5 increases trichodermin production and antimicrobial activity in *Trichoderma brevicompactum*. *Fungal genetics and biology : FG & B* **48**: 285–296.

Vens C, Rosso M-N, Danchin EGJ. **2011**. Identifying discriminative classification-based motifs in biological sequences. *Bioinformatics* **27**: 1231–1238.

Wang C-M, Hopson R, Lin X, Cane DE. **2009**. Biosynthesis of the sesquiterpene botrydial in *Botrytis cinerea*. Mechanism and stereochemistry of the enzymatic formation of presilphiperfolan-8beta-ol. *Journal of the American Chemical Society* **131**: 8360–8361.

Wawrzyn GT, Quin MB, Choudhary S, López-Gallego F, Schmidt-Dannert C. **2012**. Draft genome of *Omphalotus olearius* provides a predictive framework for sesquiterpenoid natural product biosynthesis in Basidiomycota. *Chemistry & biology* **19**: 772–783.

Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, *et al.* 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research* **43**: W237–W243.

Whittington DA, Wise ML, Urbansky M, Coates RM, Croteau RB, Christianson DW. **2002**. Bornyl diphosphate synthase: Structure and strategy for carbocation manipulation by a terpenoid cyclase.

Proceedings of the National Academy of Sciences **99**: 15375 LP – 15380.

Wu W, Tran W, Taatjes CA, Alonso-Gutierrez J, Lee TS, Gladden JM. **2016**. Rapid Discovery and Functional Characterization of Terpene Synthases from Four Endophytic *Xylariaceae*. *PLOS ONE* **11**: e0146983.

Xiao H, Zhong J-J. **2016**. Production of Useful Terpenoids by Higher-Fungus Cell Factory and Synthetic Biology Approaches. *Trends in Biotechnology* **34**: 242–255.

Zhang C, Chen X, Orban A, Shukal S, Birk F, Too H-P, Rühl M. **2020**. *Agrocybe aegerita* Serves As a Gateway for Identifying Sesquiterpene Biosynthetic Enzymes in Higher Fungi. *ACS Chemical Biology* **15**: 1268–1277.

Supporting Information

Table S1 List of the characterized fungal STS used in this study. [1] Agger et al., Mol Microbiol. 2009; [2] Wawrzyn et al., Chem Biol., 2012 ; [3] Caruthers et al., J Biol Chem, 2000; [4] Felicetti et al., J Am Chem Soc., 2004; [5] Quin et al., Chembiochem, 2013; [6] Engels et al., J Biol Chem, 2011; [7] Pinedo et al., ACS Chem Biol., 2008; [8] Wang et al., J Am Chem Soc, 2009; [9] Brock et al., Chembiochem, 2011); [10] Brock et al., Chembiochem, 2013; [11] Rynkiewicz et al., Proc Natl Acad Sci USA, 2001; [12] Cane et al., Arch Biochem Biophys., 1993; [13] McCormick et al., Appl Environ Microbiol, 2010.

	Phylum	Order	Species	Gene name	Protein Id	GenBankID	Main reaction products	Used as query in blast searches	References
Class I	Basidiomycota	Agaricales	*Coprinopsis cinereus*	Cop1		XP001832573	germacrene A	No	[1]
	Basidiomycota	Agaricales	*Coprinopsis cinereus*	Cop2		XP001836556	germacrene A	yes	[1]
	Basidiomycota	Agaricales	*Coprinopsis cinereus*	Cop3		XP001832925	a-muurolene, germacrene A	No	[1]
	Basidiomycota	Agaricales	*Omphalotus olearius*	Omp1			α-muurolene	No	[2]
	Basidiomycota	Agaricales	*Omphalotus olearius*	Omp3			a-muurolene, germacrene A	yes	[2]
	Ascomycota	Eurotiomycete	*Penicilum roqueforti*	PrAS		Q03471	aristolochene	No	[3],[4]
Class II	Basidiomycota	Russulales	*Stereum hirsutum*		128017		δ-Cadinene	yes	[5]
	Basidiomycota	Agaricales	*Coprinopsis cinereus*	Cop4		XP001836356	δ-Cadinene, cubedol.	No	[1]
	Basidiomycota	Agaricales	*Omphalotus olearius*	Omp4			δ-Cadinene	No	[2]
	Basidiomycota	Agaricales	*Omphalotus olearius*	Omp5a			epi-zonarene, γ-cadinene	yes	[2]
	Basidiomycota	Agaricales	*Omphalotus olearius*	Omp5b			γ-cadinene	No	[2]
Class III	Basidiomycota	Russulales	*Stereum hirsutum*		73029		Δ6-protoilludene	No	[5]
	Basidiomycota	Russulales	*Stereum hirsutum*		64702		Δ6-protoilludene	No	[5]
	Basidiomycota	Russulales	*Stereum hirsutum*		25180		Δ6-protoilludene	No	[5]
	Basidiomycota	Agaricales	*Omphalotus olearius*	Omp6			Δ6-protoilludene	yes	[2]
	Basidiomycota	Agaricales	*Omphalotus olearius*	Omp7			Δ6-protoilludene	No	[2]
	Basidiomycota	Agaricales	*Coprinopsis cinereus*	Cop5		XP001834007	pentalenene	No	[1]
	Basidiomycota	Agaricales	*Armilaria gallica*	Armga1		AGR34199	Δ6-protoilludene	No	[6]
	Ascomycota	Leotiomycete	*Botrytis cinerea*	BcBOT2		AAQ16575	presilphiperfolan-8ß-ol	yes	[7]; [8]
	Ascomycota	Sordariomycete	*Fusarium fujikuroi*	Ffsc4		CCP20071	koraiol	No	[9]; [10]
Class IV	Basidiomycota	Russulales	*Stereum hirsutum*		159379		β-barbatene	No	[5]
	Basidiomycota	Agaricales	Coprinopsis cinereus	Cop6		XP001832548	α-cupranene	No	[1]
	Basidiomycota	Agaricales	*Omphalotus olearius*	Omp9			α/β-barbatene	No	[2]
	Basidiomycota	Agaricales	*Omphalotus olearius*	Omp10			daucene/trans-dauca-4(11),8-diene	yes	[2]
	Basidiomycota	Polyporales	*Fomitopsis pinicola*	Fompi84944			α-cuprenene	yes	[2]
	Ascomycota	Sordariomycete	*Fusarium fujikuroi*	Ffsc6		CCP20072	(-)-α-acorenol	yes	[9]; [10]
	Ascomycota	Sordariomycete	*Fusarium spordrichoides*	FsTS		AAN05035	trichodiene	No	[11], [12]
	Ascomycota	Sordariomycete	*Fusarium graminearum*	FgCLM1		ACY69978	longiborneol	No	[13]

Table S2 List of the 24 Polyporales genomes used in this study.

Species	Genome
Abortiporus biennis	Abobie1
Artolenzites elegans (1663)	Artel1
Earliella scabrosa	Eaesca1
Fomes fomentarius	Fomfom1
Rhodofomes roseus	Fomros1
Hexagonia nitida	Hexnit1
Irpex lacteus	Irplac1
Leiotrametes sp.	Leisp1
Fomitopsis betulina	Pipbet1
Polyporus brumalis	Polbr1
Pycnoporus cinnabarinus (137)	Pycci1
Pycnoporus coccineus (310)	Pycco1
Pycnoporus puniceus	Pycpun1
Pycnoporus sanguineus	Pycsa1
Trametes betulina	Trabet1
Trametopsis cervina	Trace1
Trametes cingulata	Traci1
Trametes gibbosa	Tragib1
Trametes lactinea	Tralac1
Trametes ljubarskyi	Tralj1
Trametes maxima	Tramax1
Trametes menziesii	Tramen1
Trametes meyenii	Tramey1
Coriolopsis polyzona	Trapol1

Table S3 Percentage of Identity between selected plant and fungal STS C-term part of the sequences

		Fungi								
		Cop5	Stehi_73029	Omp7	Ompol1	Stehi_128017	Ompol1_0	Omp4	Cop2	Omp1
Plants	A4FVP2|TPS03	24%	20%	22%	19%	24%	22%	23%	23%	25%
	Q9ZUH4|TPSA_ARATH	25%	23%	23%	21%	26%	23%	24%	23%	28%
	B9T536|TPS10_RICCO	26%	23%	23%	21%	24%	21%	23%	24%	26%
	Q5UB07|TPS4_MEDTR	28%	22%	23%	22%	24%	23%	23%	23%	26%
	E3W207|SAUSS_SANAS	27%	23%	25%	21%	24%	23%	23%	23%	28%
	F6M8H4|SAST_SANAL	27%	23%	24%	21%	23%	23%	23%	23%	27%
	E3W208|SPISS_SANSP	26%	23%	24%	21%	23%	23%	23%	23%	28%
	Q4KSH9|BARS_ARATH	22%	20%	22%	18%	20%	20%	21%	22%	25%
	Q9LIA1|TPS25_ARATH	26%	22%	23%	21%	23%	23%	23%	23%	27%
	Q6JD70|TPS5A_MAIZE	27%	23%	25%	22%	26%	26%	25%	24%	25%

Table S4 Assignment of characterized STS to each class, based on 1st, 2nd and 3rd generation of HMM models. Wrong assignments as compared to C. Schmidt-Dannert et al. (2014) are highlighted in bold.

STS	JGI ID	expected classification	Classification based on 1st generation HMM profiles	Classification based on 2nd generation HMM profiles
Cop2	jgi|Copci_AmutBmut1|	Class 1	class1	class1
Cop1	jgi|Copci1|3117		class1	class1
Omp1	jgi|Ompol1|1311		class1	class1
Cop3	jgi|Copci_AmutBmut1|		class1	class1
Omp3	jgi|Ompol1|4636		class1	class1
Omp5a	jgi|Ompol1|2392	Class 2	class2	class2
Omp5b	jgi|Ompol1|2393		class2	class2
Omp4	jgi|Ompol1|1447		class2	class2
Cop4	jgi|Copci_AmutBmut1|		class2	class2
Stehi	jgi|Stehi1|128017		class2	class2
Ffsc4	jgi|Fusfu1|11322	Class 3	**class 1**	**class 1**
Stehi 73029	jgi|Stehi1|73029		**class2**	class3
Armga1	jgi|Armga1|1000221		**class1**	class3
Stehi 64702	jgi|Stehi1|64702		class3	class3
Omp7	jgi|Ompol1|2271		class3	class3
Stehi 25180	jgi|Stehi1|25180		class3	class3
Omp6	jgi|Ompol1|4774		**class1**	class3
Ffsc6	jgi|Fusfu1|12876	Class 4	**class-unknown**	**NA**
Fompi	jgi|Fompi3|84944		class4	class4
Stehi	jgi|Stehi1|159379		class4	class4
Cop6	jgi|Copci1|3094		class4	class4
Omp9	jgi|Ompol1|3258		class4	class4
Omp10	jgi|Ompol1|3981		class4	class4

Chapter V CONCLUSIONS

Adaptation is an important process shaping fungal species diversity (Ellison *et al.*, 2011; Branco *et al.*, 2017; Robin *et al.*, 2017). It is translated by evolutionary histories, such as the evolution of advantageous traits, that determine the ability of a species to colonize and withstand a certain habitat (Branco S., 2019). The sequencing of a large number of fungal genomes has made it possible to reveal a wide range of genomic changes, such as genomic rearrangements, gene gains, gene losses or introgressions. Gene duplication allows functional diversification while gene deletion allows the elimination of unnecessary enzymatic machineries to save energy and, in the case of pathogens, to avoid detection by host defences (Gladieux *et al.*, 2014). Such evolutionary changes play big roles in population polymorphism or even speciation (Sobel *et al.*, 2010; Nosil P., 2012).

In this thesis, we aimed to access the genomic and functional diversity among wood-decay fungi at different taxonomic levels and to explore the genomic basis of evolutionary adaptations to the wood-decay lifestyle.

For this purpose, we started by comparing the genomes, transcriptomes and secretomes of four species from the genus *Pycnoporus,* which are encountered in different geoclimatic areas. We observed a strong intra-genus genomic conservation. Indeed, the four *Pycnoporus* species shared similar genomic features, including similar genome size (30 to 36 Megabases) and absence of major genome rearrangements, possibly related to their low coverage in transposable elements (1.8% to 12.3%). In addition, they share a high proportion of conserved protein coding genes (~82%) and possess a low proportion of species-specific genes (4 to 5%). Finally, they do not show remarkable differences in their CAZyme gene repertoires, in which the LPMOs, Class II PODs and GMC-oxidoreductase genes are abundant, a common feature of white-rot fungi. All those findings suggest that 39 million years of evolution and geographical isolation of the species has not yet led to any noticeable genome level diversity in the genus *Pycnoporus*.

This somewhat surprising finding suggested wood decayers do not share the same evolutionary trends as pathogens, which face an ”arms race” situation in which the fungi actively rearrange their genome to hijack ever-evolving host defences. Similarly, some mycorrhizal fungi showed strong abilities to conquer large host ranges. For example, isolates of the arbuscular mycorrhizal *Rhizophagus irregularis* harbour a large number of isolate-specific genes and a large variability in their gene repertoires and TE content. It is proposed that this striking genome diversity played a role in their adaptation to fluctuating environments and their ability to establish symbiotic interactions with a wide range of plant hosts (Chen *et al.*, 2018). In addition, it has been shown that the transition from saprotrophy to ectomycorrhizal symbiosis included, among others, the acquisition of small secreted proteins (SSPs) and the proliferation

of TEs (Miyauchi *et al.*, 2020). Those latter could play a key role in promoting the rapid evolution of SSPs that may facilitate the accommodation of symbiotic fungi within a wide range of host plants. Similarly, in biotrophic pathogens, such as *Blumeria graminis* f. sp. *hordei* and *B. graminis* f. sp. *tritici*, it has been proposed that bursts of TEs offer a template for rapid evolution of virulence genes located in TE-rich regions, by causing their duplications, rearrangements, or deletions (Frantzeskakis *et al.*, 2018). As another example, the comparison of six French *Fusarium graminearum* isolates revealed a remarkable number of single nucleotide polymorphisms in genes involved in adaptation to biotic and abiotic stress (Laurent *et al.*, 2017). Strikingly, the less toxinogenic and less aggressive isolates showed accumulation of non-synonymous mutations in the genes involved in trichothecene mycotoxin synthesis. Hence in pathogens and symbionts, the genomic plasticity allows aggressiveness and virulence variation and promotes adaptation to the hosts.

The observed genomic conservation among wood decayers from the genus *Pycnoporus* was comforted by our extended study on the Polyporales order, in which 26 newly sequenced genomes were included and compared along with 24 previously published ones. The 50 genomes showed relatively low TE coverage (0.37% to 16%). We observed reduced gene counts for plant cell wall degrading CAZymes in brown-rotters (all from the antrodia clade), as expected from the gene loss events that accompanied the transition from white-rot toward brown-rot decay mode (Floudas *et al.*, 2012). These reductions affected laccase, AA9 LPMO, copper radical oxidase (AA5_1), expansins, PL14 pectate lyase and CBM1 gene counts. Interestingly however, we observed that the species from the sister clade gelatoporia, which are white-rotters, had intermediate gene counts for AA9 LPMOs, expansins and pectate lyases, suggesting that gene losses already occurred in the most recent common ancestor (MRCA) of antrodia and gelatoporia, followed by further loss events in the antrodia clade. Alternatively, gene expansions could have occurred later on in the gelatoporia lineage, as showed by our analysis for Class II peroxidases active on lignin.

In our study, species having the same decay mode, notably the brown rotters on the one hand or white rotters on the other hand, didn't show significative differences in their CAZyme gene repertoires. The overall conservation of CAZomes in Polyporales could be explained by the already diverse lignocellulosic toolkit present in the common ancestor of Polyporales. Indeed, white-rot decay most likely arouse in the Carboniferous period in the MRCA of Auriculariales and higher Agaricomycetes, which predates the Polyporales MRCA by at least 100-150 million years (Floudas *et al.*, 2012; Varga *et al.*, 2019). This may have provided early members of the Polyporales with the necessary tools to degrade plant cell walls.

This observed conservation in repertoires of protein coding genes among Polyporales might be partly related to the dead wood habitat. Regarding the occurrence of Polyporales species on

preferential woods, some authors indicate that brown rotters are preferentially encountered on wood from gymnosperm trees and white rotters on that from angiosperms. However, Krah *et al.* (2018) revealed that high transition rates between host-specialists and generalists have occurred in both decay modes, which confers overall generalist abilities to grow on both wood types.

A comparative analysis of Polyporales with 57 other Agaricomycetes (done during my PhD) highlighted no peculiarities in the gene repertoires of CAZymes, cytochrome P450 and transporters (ABC and MFS) between wood-decayers in Agaricomycetes. This enlarged analysis supported our hypothesis that the capacities to degrade the plant cell wall and to detoxify the toxic products released during wood degradation were acquired before the diversification of Polyporales.

Surprisingly, within the order Polyporales, we observed different expansion/contraction trajectories for CAZyme genes among the white-rot clades, which eventually generated similar gene counts in extant species. Indeed, large differences were observed in the number of gene duplications inferred in the MRCA of the core polyporoid vs that of the phlebioid clade (103 duplications vs nine), notably for GMC-oxidoreductases, expansins and CBM1-containing genes. We speculate that further gene family expansions in the phlebioid lineage eventually resulted in similar gene counts for these CAZymes in extant core polyporoid and phlebioid species, as showed here for Class II peroxidase genes. Only in the case of laccases, the phlebioid species showed a reduced number of genes as compared to other Polyporales clades.

Despite wood decayers are mostly evolving to become host-generalists as explained above, they still show some polymorphism in their abilities to grow on lignocellulosic substrates and in the enzymatic systems they mobilize. In the case of class II peroxidases for example, the diverse expansion trajectories support functional diversity. Indeed, we observed important recent gene gain events for manganese peroxidases (MnP) in core polyporoid lineages (leading to *Trametes versicolor*, *Hexagonia nitida, Dichomitus squalens*), in gelatoporia lineages (*Obba rivulosa*, *Ceriporiopsis subvermispora*), phlebioid lineages (*Phlebia brevispora*, *Phanerochaete carnosa*), and in residual polyporoid lineages (*Abortiporus biennis, Panus rudis*). The versatile peroxidases (VP) have undergone abundant recent gene gains in the core polyporoid and residual polyporoid lineages that lead to *Polyporus* spp., *Lentinus tigrinus*, *Fomes fomentarius*, *Earliella scabrosa*, *Hexagonia nitida* and *Panus rudis*. Finally, lignin peroxidases (LiP) underwent expansions in some Trametoid and in phlebioid lineages leading to *Phanerochaete chrysosporium*, *Bjerkandera adusta* and *Phlebia brevispora*. These findings are in line with the selection by different research groups of fungi of biotechnological interest, that have been studied extensively for their ability to degrade lignocellulose (e.g. *Trametes versicolor, Dichomitus squalens, Phlebia brevispora, Phanerochaete carnosa, Phanerochaete chrysosporium*, *Bjerkandera adusta),* and enlighten the recent discovery of *Polyporus brumalis* as an efficient degrader of lignin (Miyauchi et al., 2018).

Interestingly, they shed light on yet understudied species that could be of great interest to biotechnologies aimed at plant biomass conversion (e.g. *Hexagonia nitida, Earliella scabrosa, Obba rivulosa, Panus rudis).*

Interestingly, we also observed gene expansion in early Polyporales for several groups of regulator families for which we do not have evidence yet of their involvement in wood decay, including candidate protein kinases, candidate transcription factors, and proteins with F-box or Leucine-rich repeat kinase-like (LLR) domains. This led to the hypothesis that despite the ancestors of the Polyporales already possessed a strong genetic material for lignocellulose degradation, the diversification of some regulatory families that may be functionally linked to wood decay followed only after the origin of the order.

The proportion of species-specific genes in the Polyporales genomes varied from 11 to 43%, with an average of 80% annotated as SSPs. As presented above, some of the characterized SSPs are effectors likely required for symbiosis or parasitism development to impair the defense response of the host. The consistent SSP repertoires we highlighted in Polyporales, suggest a new possible role of SSPs in saprotrophic lifestyles. Previous observations on the genomes of saprotroph fungi, including white rotters and brown rotters, already showed that they encode numerous SSPs (Pellegrin *et al.*, 2015; Kim *et al.*, 2016), which could be used as weapons or decoys in intra- and interspecific cell-to-cell communication with other microorganisms sharing the same ecological niches or as regulators of quorum sensing (Feldman *et al.*, 2020). The fact that most of the SSPs in Polyporales seem to be species-specific, open the possibility that some species might transition towards host-specificity rather than being generalist.

In laboratory conditions, different fungal species show different abilities to grow on diverse lignocellulosic substrates. This phenotype diversity highlights functional diversity beyond the relative conservation of gene repertoires. Within the genus *Pycnoporus,* this polymorphism was related with different gene transcription profiles between the species. This finding prove that gene repertoires obtained from genome sequencing are not sufficient to predict the ability of a strain to grow on recalcitrant raw biomass and that transcriptomic analyses are necessary to understand the enzymatic systems at play for lignocellulose degradation. A comparative transcriptomic and secretomic analysis allowed us to identify a set of CAZymes commonly present in the secretomes of *Pycnoporus* species and sharing similar transcriptional regulations in response to cellulosic or lignocellulosic substrates. These included genes coding for enzymes active on; β-1,4-glucan linkages (GH5_5, GH6, GH7, GH131 and AA9 LPMOs); the hemicellulose backbone (GH5_7, GH10, GH12, GH16, GH27, GH51 and GH74) or branched groups (CE1); pectin (GH28) and lignin (AA1_1). Hence, this set of genus-wide conserved genes represents functional markers for substrate decomposition activity by these fungi. Furthermore, we comforted the set of genes that represent a core response to lignocellulose by comparing the transcriptomes of six white-rot species from the order Polyporales. This transcriptomic comparison highlighted 11 genes, out of which eight were already found in the conserved set of the *Pycnoporus* genus

(GH5_5, GH6, GH7, GH10, GH131, GH28, CE1 and AA9) and three were new candidates (GH1, GH3 and CE16). In addition, this expended analysis highlighted AA9 genes appended to an atypical ~30 amino acid module commonly up-regulated in the six species. The X282 module is to date only found in Agaricomycetes. Further studies are required to understand its possible contribution to wood decay. Apart from CAZymes, SSP genes with a thaumatin domain, candidate lipases and cupredoxin proteins were found to be commonly up-regulated in at least five of the six Polyporales species analyzed, suggesting they might be new candidate players in wood decomposition.

Apart from the diverse enzymatic mechanisms to degrade the complex lignocellulosic biopolymers, fungi have evolved strategies to communicate with other microorganisms and defend themselves against biotic and abiotic stress in their natural habitats. These strategies are reflected by the production of a wide range of secondary metabolites such as terpene volatile compounds. The tool we built for the genome mining of sesquiterpene synthases (STS) in the fungal kingdom allowed us to explore the taxonomical diversity of such enzymes. Interestingly, we found that, unlike CAZymes, STS gene repertoires were highly different among Polyporales (ranging from 4 to 20 copies). This observed STS diversity might be a result of Polyporales adaptation to different stress levels or different microbial communities in their ecological niches. Indeed, Polyporales face constant competition with other microorganisms for the available resources. This observation may explain the higher diversity of terpene synthases as compared to lignocellulose degrading enzymes among wood decayers. In addition, we highlighted differences at the phylum level, where we showed that basidiomycetes encode 2 to 3 times more STS genes compared to ascomycetes. Our preliminary study suggested that this difference could be related to the higher level of tandem duplication events that occurred in basidiomycetes as compared to ascomycetes. It has been shown in previous studies that Basidiomycota rely mostly on terpenoids as their predominant class of secondary metabolites, while Ascomycota have evolved a different arsenal, with polyketides and non-ribosomal peptides as major classes (Boettger & Hertweck, 2013). The observed taxonomical differences in STS gene distribution between species from these two Dikarya phyla and within the Polyporales order suggest that fungal terpene compounds might have a wide functional diversity. Using our newly developed tool, we could predict the mechanism of cyclisation of farnesyl pyrophosphate by some of the identified STS based on sequence similarity with previously characterized STS. Both in basidiomycetes and ascomycetes, STS that are predicted to catalyze the production of the bisabolol cation (via 1,6 closure of (3R)-nerolidyl diphosphate; herein called class 4 enzymes), are the most abundant. This finding was congruent with the recent identification of abundant bisabolol-derived terpenes in the volatile compounds produced by *Schizophyllum commune* in competition assays with other fungi (Wirth et al., 2021). Among the Polyporales genomes, the differences in total STS gene counts are mostly caused by fluctuations in gene

counts for class 4 and class 3 enzymes (which catalyze the formation of trans-humulyl cation via 1,11 closure of 2E,6E-FPP). Thus, we anticipate that Polyporales diversity in the STS compounds they produce is largely related to the diversity of enzymes with mechanism of cyclisation similar to those of clades 3 and 4. Yet, exploring the functional diversity of the newly identified STS will require further analyses on their sequence polymorphism, in combination with biochemical assays to identify the terpene backbones produced by the enzymes as well as biological activity assays to test, for example, their potential role as antioxidant or anti-microbial compounds. Finally, the proper identification of STS genes in fungi opens the possibility to mine the genomes for biosynthesis gene clusters involved in the production of decorated terpenes with additional biological activities.

Chapter VI PERSPECTIVES

Following our exploratory study on the genomes of wood-decay fungi, it could be interesting to correlate genomic information with ecological ones, to further explain from a genomic view point, the succession of saprotrophic species in their ecological niches. In fact, it is well documented that saprotrophic fungi have a hierarchical community organization related to their competitive ability, with some species co-occurring at a very specific decay stage depending on predecessors, while others can utilize wood components at a broader time interval. Theoretically, primary colonizers are pioneer species with high breeding efficiencies, secondary colonizers follow as competitor species with an ability to decompose recalcitrant lignocellulose and finally superior colonizers are species tolerant to stress that benefit from the available substrates modified by predecessors. Consequently, the present comparative genomics on wood-decay fungi along with the knowledge on each fungus ability to decompose lignocellulose, to be good competitors or to be stress-tolerant, could be a good basis for the study of the genetic differences between primary, secondary or late-stage saprotrophic decomposers. Given the central role of saprotrophic fungi in wood decomposition, understanding the functional traits underlying their dynamics is a first step towards untangling the wood decomposition processes in nature, which could further lead to better comprehend the process of carbon release in forest ecosystems.

From a biotechnological perspective, our analysis on the evolutionary histories of PCWDE and their diversity among Polyporales could help targeting species with optimal enzyme portfolios for lignocellulose degradation. For example, we identified species showing expansion of different class II peroxidase types, such as manganese peroxidases in yet overlooked species (e.g. *Hexagonia nitida*, *Obba rivulosa*, *Panus rudis*), versatile peroxidases in core polyporoid species (e.g. *Earliella scabrosa*, *Fomes fomentarius* and *Polyporus brumalis*) or lignin peroxidase gene gains in phlebioid species (e.g. *Phanerochaete chrysosporium* and *Bjerkandera adusta*). Also, in a previous study, we highlighted that the extraordinary ability of *P. brumalis* for lignin degradation was related to the up-regulation, secretion and expansion of an abundant set of versatile peroxidases and manganese peroxidases. Therefore, along with other studies focusing on different enzymes, we enlarged the range of attractive fungal species that could be further screened for a particular industrial application. Finally, the conserved sets of enzymes produced by different species during growth on diverse lignocellulosic substrates could be used to improve enzymatic cocktails dedicated to biomass degradation by targeting all PCW components.

Following the identification of sesquiterpene cyclases across the fungal kingdom, biochemical analyses will be required to validate their activities as cyclases of the linear farnesyl diphosphate precursor. Furthermore, the structural characterization of the produced terpenes will be needed to validate the predicted mechanisms of cyclisation of the newly identified fungal STSs.

Afterwards, interesting analyses could be done to explore the biosynthesis gene clusters of the identified STS, i.e. the suites of biosynthesis genes co-located in close vicinity in the genomes. Indeed, the diversity of terpene compounds produced by fungi is not only related to the large variety of their hydrocarbon backbones, but also to the addition of functional groups by tailoring enzymes that may change dramatically the chemo-physical nature of the resulting terpenoids. The identification of those terpene tailoring enzymes can help to enlarge the library of characterized terpenes produced by the fungi, mainly those that are difficult to produce biochemically.

In addition, when validated, the tool we developed could greatly accelerate the identification of the STS enzymes of interest for terpenoid production via synthetic biology. Indeed, the tool will facilitate the identification of the terpenes and terpenoids involved in microbial communications during wood decay in natural environments. The exploration of the diversity of terpene chemical structures accessible from the fungi, could also help in the discovery of new terpenes and terpenoids with biological activities of high industrial values such as antioxidant, anti-microbial, insect attractant/repellant...

REFERENCES

Agger S, Lopez-Gallego F, Schmidt-Dannert C. **2009**. Diversity of sesquiterpene synthases in the basidiomycete *Coprinus cinereus*. *Molecular Microbiology* **72**: 1181–1195.

Alfaro M, Majcherczyk A, Kües U, Ramírez L, Pisabarro AG. **2020**. Glucose counteracts wood-dependent induction of lignocellulolytic enzyme secretion in monokaryon and dikaryon submerged cultures of the white-rot basidiomycete *Pleurotus ostreatus*. *Scientific Reports* **10**: 12421.

Amore A, Giacobbe S, Faraco V. **2013**. Regulation of cellulase and hemicellulase gene expression in fungi. *Current genomics* **14**: 230–249.

Anasontzis GE, Lebrun M-H, Haon M, Champion C, Kohler A, Lenfant N, Martin F, O'Connell RJ, Riley R, Grigoriev I V, *et al.* 2019. Broad-specificity GH131 β-glucanases are a hallmark of fungi and oomycetes that colonize plants. *Environmental Microbiology* **21**: 2724–2739.

Anderson JB, Stasovski E. **1992**. Molecular Phylogeny of Northern Hemisphere Species of *Armillaria*. *Mycologia* **84**: 505–516.

Andlar M, Rezić T, Marđetko N, Kracher D, Ludwig R, Šantek B. **2018**. Lignocellulose degradation: An overview of fungi and fungal enzymes involved in lignocellulose degradation. *Engineering in Life Sciences* **18**: 768–778.

Andrew M, Barua R, Short SM, Kohn LM. **2012**. Evidence for a Common Toolbox Based on Necrotrophy in a Fungal Lineage Spanning Necrotrophs, Biotrophs, Endophytes, Host Generalists and Specialists. *PLOS ONE* **7**: e29943.

El Ariebi N, Hiscox J, Scriven SA, Müller CT, Boddy L. **2016**. Production and effects of volatile organic compounds during interspecific interactions. *Fungal Ecology* **20**: 144–154.

Ayuso-Fernández I, Ruiz-Dueñas FJ, Martínez AT. **2018**. Evolutionary convergence in lignin-degrading enzymes. *Proceedings of the National Academy of Sciences* **115**: 6428 LP – 6433.

Beckham GT, Matthews JF, Peters B, Bomble YJ, Himmel ME, Crowley MF. **2011**. Molecular-Level Origins of Biomass Recalcitrance: Decrystallization Free Energies for Four Common Cellulose Polymorphs. *The Journal of Physical Chemistry B* **115**: 4118–4127.

Begoude BAD, Slippers B, Wingfield MJ, Roux J. **2011**. The pathogenic potential of endophytic *Botryosphaeriaceous* fungi on *Terminalia* species in Cameroon. *Forest Pathology* **41**: 281–292.

Berbee ML, James TY, Strullu-Derrien C. **2017**. Early Diverging Fungi: Diversity and Impact at the Dawn of Terrestrial Life. *Annual Review of Microbiology* **71**: 41–60.

Berrin J-G, Navarro D, Couturier M, Olivé C, Grisel S, Haon M, Taussac S, Lechat C, Courtecuisse R, Favel A, *et al.* 2012. Exploring the natural fungal biodiversity of tropical and temperate forests toward improvement of biomass conversion. *Applied and environmental microbiology* **78**: 6483–6490.

Berrin J-G, Rosso M-N, Abou Hachem M. **2017**. Fungal secretomics to probe the biological functions of lytic polysaccharide monooxygenases. *Carbohydrate research* **448**: 155–160.

Bissaro B, Røhr ÅK, Müller G, Chylenski P, Skaugen M, Forsberg Z, Horn SJ, Vaaje-Kolstad G, Eijsink VGH. **2017**. Oxidative cleavage of polysaccharides by monocopper enzymes depends on H2O2. *Nat Chem Biol.* **13**: 1123.

Blackwell M. **2011**. The fungi: 1, 2, 3 ... 5.1 million species? *American journal of botany* **98**: 426–438.

Blanchette R.A. B. 1984. Manganese accumulation in wood decayed by white rot fungi. *Phytopathology* **74**:725-730

Blanchette RA. **1991**. Delignification by wood-decay fungi. *Annu Rev Phytopathol.* **29**: 381–403.

Boerjan W, Ralph J, Baucher M. **2003**. Lignin Biosynthesis. *Annual Review of Plant Biology* **54**: 519–546.

Boettger D, Hertweck C. **2013**. Molecular diversity sculpted by fungal PKS-NRPS hybrids. *Chembiochem : a European journal of chemical biology* **14**: 28–42.

Branco S. **2019**. Fungal diversity from communities to genes. *Fungal Biology Reviews* **33**: 225–237.

Branco S, Bi K, Liao H-L, Gladieux P, Badouin H, Ellison CE, Nguyen NH, Vilgalys R, Peay KG, Taylor JW, *et al.* 2017. Continental-level population differentiation and environmental adaptation in the mushroom *Suillus brevipes*. *Molecular Ecology* **26**: 2063–2076.

van den Brink J, de Vries RP. **2011**. Fungal enzyme sets for plant polysaccharide degradation. *Applied microbiology and biotechnology* **91**: 1477–1492.

Brundrett M. **2004**. Diversity and classification of mycorrhizal associations. *Biological Reviews* **79**: 473–495.

Buchvaldt Amby D, Manczak T, Petersen MA, Sundelin T, Weitzel C, Grajewski M, Simonsen HT, Jensen B. **2016**. Role of the *Colletotrichum acutatum* sesquiterpene synthase CaTPS in the biosynthesis of sesquiterpenoids. *Microbiology (Reading, England)* **162**: 1773–1783.

Buckling A, Wills MA, Colegrave N. **2003**. Adaptation Limits Diversification of Experimental Bacterial Populations. *Science* **302**: 2107 LP – 2109.

Buis A. **2019**. *The Atmosphere: Getting a Handle on Carbon Dioxide*.

Campbell MM, Sederoff RR. **1996**. Variation in Lignin Content and Composition (Mechanisms of Control and Implications for the Genetic Improvement of Plants). *Plant physiology* **110**: 3–13.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. **2009**. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* **37**: D233–D238.

Carroll G. **1988**. Fungal Endophytes in Stems and Leaves: From Latent Pathogen to Mutualistic Symbiont. *Ecology* **69**: 2–9.

Chai R, Qiu C, Liu D, Qi Y, Gao Y, Shen J, Qiu L. **2013**. β-Glucan Synthase Gene Overexpression and β-Glucans Overproduction in *Pleurotus ostreatus* Using Promoter Swapping. *PLOS ONE* **8**: e61693.

Chang Y, Wang S, Sekimoto S, Aerts AL, Choi C, Clum A, LaButti KM, Lindquist EA, Yee Ngan C, Ohm RA, *et al.* 2015. Phylogenomic Analyses Indicate that Early Fungi Evolved Digesting Cell Walls of Algal Ancestors of Land Plants. *Genome Biology and Evolution* **7**: 1590–1601.

Cheeseman K, Ropars J, Renault P, Dupont J, Gouzy J, Branca A, Abraham A-L, Ceppi M, Conseiller E, Debuchy R, *et al.* 2014. Multiple recent horizontal transfers of a large genomic region in cheese making fungi. *Nature communications* **5**: 2876.

Chen M, Al-lami N, Janvier M, D'Antonio EL, Faraldos JA, Cane DE, Allemann RK, Christianson DW. **2013**. Mechanistic Insights from the Binding of Substrate and Carbocation Intermediate Analogues to Aristolochene Synthase. *Biochemistry* **52**: 5441–5453.

Chen ECH, Morin E, Beaudet D, Noel J, Yildirir G, Ndikumana S, Charron P, St-Onge C, Giorgi J, Krüger M, *et al.* 2018. High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. *The New phytologist* **220**: 1161–1171.

Christianson DW. **2008**. Unearthing the roots of the terpenome. *Current Opinion in Chemical Biology* **12**: 141–150.

Couturier M, Ladevèze S, Sulzenbacher G, Ciano L, Fanuel M, Moreau C, Villares A, Cathala B, Chaspoul F, Frandsen KE, *et al.* 2018. Lytic xylan oxidases from wood-decay fungi unlock biomass degradation. *Nature chemical biology* **14**: 306–310.

Cowling EB. **1961**. Comparative biochemistry of the decay of sweetgum sapwood by white-rot and brown-rot fungi. United States Department of Agriculture.

Daly P, López SC, Peng M, Lancefield CS, Purvine SO, Kim Y-M, Zink EM, Dohnalkova A, Singan VR, Lipzen A, *et al.* 2018. *Dichomitus squalens* partially tailors its molecular responses to the composition of solid wood. *Environ Microbiol.* **20**: 4141–4156.

Daly P, Peng M, Di Falco M, Lipzen A, Wang M, Ng V, Grigoriev I V, Tsang A, Mäkelä MR, de Vries RP. **2019**. Glucose-mediated repression of plant biomass utilization in the white-rot fungus *Dichomitus squalens*. *Applied and Environmental Microbiology*.

Dashtban M, Maki M, Leung KT, Mao C, Qin W. **2010**. Cellulase activities in biomass conversion: measurement methods and comparison. *Critical Reviews in Biotechnology* **30**: 302–309.

Ding S-Y, Himmel ME. **2006**. The Maize Primary Cell Wall Microfibril: A New Model Derived from Direct Visualization. *Journal of Agricultural and Food Chemistry* **54**: 597–606.

Douzery EJP, Snell EA, Bapteste E, Delsuc F, Philippe H. **2004**. The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? *Proceedings of the National Academy of Sciences* **101**: 15386–15391.

Drew DP, Andersen TB, Sweetman C, Møller BL, Ford C, Simonsen HT. **2016**. Two key polymorphisms in a newly discovered allele of the Vitis vinifera TPS24 gene are responsible for the production of the rotundone precursor α-guaiene. *Journal of experimental botany* **67**: 799–808.

Durairaj J, Di Girolamo A, Bouwmeester HJ, de Ridder D, Beekwilder J, van Dijk ADJ. **2019**. An analysis of characterized plant sesquiterpene synthases. *Phytochemistry* **158**: 157–165.

E4tech (UK) Ltd for LBNet. **2017**. *UK Top Bio-based Chemicals Opportunities*.

Ellison CE, Hall C, Kowbel D, Welch J, Brem RB, Glass NL, Taylor JW. **2011**. Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *Proceedings of the National Academy of Sciences* **108**: 2831 LP – 2836.

Engels B, Heinig U, Grothe T, Stadler M, Jennewein S. **2011**. Cloning and characterization of an *Armillaria gallica* cDNA encoding protoilludene synthase, which catalyzes the first committed step in the synthesis of antimicrobial melleolides. *The Journal of biological chemistry* **286**: 6871–6878.

European Commission. 2012. *Innovating for sustainable growth: A bioeconomy for Europe*.

Feldman D, Yarden O, Hadar Y. **2020**. Seeking the Roles for Fungal Small-Secreted Proteins in Affecting Saprophytic Lifestyles. *Frontiers in Microbiology* **11**: 455.

Filiatrault-Chastel C, Navarro D, Haon M, Grisel S, Herpoël-Gimbert I, Chevret D, Fanuel M, Henrissat B, Heiss-Blanquet S, Margeot A, *et al.* 2019. AA16, a new lytic polysaccharide monooxygenase family identified in fungal secretomes. *Biotechnology for Biofuels* **12**: 55.

Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martínez AT, Otillar R, Spatafora JW, Yadav JS, *et al.* 2012. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* **336**: 1715–1719.

Floudas D, Held BW, Riley R, Nagy LG, Koehler G, Ransdell AS, Younus H, Chow J, Chiniquy J, Lipzen

A, *et al.* 2015. Evolution of novel wood decay mechanisms in Agaricales revealed by the genome sequences of *Fistulina hepatica* and *Cylindrobasidium torrendii*. *Fungal genetics and biology : FG & B* **76**: 78–92.

Fong M, Berrin J-G, Paës G. **2016**. Investigation of the binding properties of a multi-modular GH45 cellulase using bioinspired model assemblies. *Biotechnology for Biofuels* **9**: 12.

Frantzeskakis L, Kracher B, Kusch S, Yoshikawa-Maekawa M, Bauer S, Pedersen C, Spanu PD, Maekawa T, Schulze-Lefert P, Panstruga R. **2018**. Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen. *BMC Genomics* **19**: 381.

Fridahl M, Lehtveer M. **2018**. Bioenergy with carbon capture and storage (BECCS): Global potential, investment preferences, and deployment barriers. *Energy Research & Social Science* **42**: 155–165.

Gilbertson RL. **1980**. Wood-rotting fungi of North America. *Mycologia*.

Gladieux P, Ropars J, Badouin H, Branca A, Aguileta G, de Vienne DM, Rodríguez de la Vega RC, Branco S, Giraud T. **2014**. Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Molecular Ecology* **23**: 753–773.

Gold MH, Youngs HL, Gelpke MD. **2000**. Manganese peroxidase. *Met Ions Biol Syst* **37**: 559–586.

Goodell B, Qian Y, Jellison J. **2008**. Fungal Decay of Wood: Soft Rot—Brown Rot—White Rot. In: ACS Symposium Series. Development of Commercial Wood Preservatives. American Chemical Society, 2–9.

Gostinčar C, Grube M, De Hoog S, Zalar P, Gunde-Cimerman N. **2009**. Extremotolerance in fungi: evolution on the edge. *FEMS Microbiology Ecology* **71**: 2–11.

Grigoriev I V, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, *et al.* 2014. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* **42**: D699–D704.

Gruber S, Seidl-Seiboth V. **2012**. Self versus non-self: fungal cell wall degradation in Trichoderma. *Microbiology* **158**: 26–34.

Gupta VK, Steindorff AS, de Paula RG, Silva-Rocha R, Mach-Aigner AR, Mach RL, Silva RN. **2016**. The Post-genomic Era of *Trichoderma reesei*: What's Next? *Trends in Biotechnology* **34**: 970–982.

Heckman DS, Geiser DM, Eidell BR, Stauffer RL, Kardos NL, Hedges SB. **2001**. Molecular Evidence for the Early Colonization of Land by Fungi and Plants. *Science* **293**: 1129–1133.

Henriksson G, Johansson G, Pettersson G. **2000**. A critical review of cellobiose dehydrogenases. *Journal of biotechnology* **78**: 93–113.

Henrissat B, Davies G. **1997**. Structural and sequence-based classification of glycoside hydrolases. *Current Opinion in Structural Biology* **7**: 637–644.

Higuchi T. **1990**. Lignin biochemistry: biosynthesis and biodegradation. *Wood Science and Technology*.

Himmel ME, Baker JO, Overend RP. **1994**. Enzymatic conversion of biomass for fuels production. *American Chemical Society,*.

Hiscox J, O'Leary J, Boddy L. **2018**. Fungus wars: basidiomycete battles in wood decay. *Studies in mycology* **89**: 117–124.

Hoegh-Guldberg O, Jacob D, Taylor M, Bindi M, Brown S, Camilloni I, Diedhiou A, Djalante R, Ebi KL, Engelbrecht F, *et al.* 2018. Impacts of 1.5°C Global Warming on Natural and Human Systems. *IPCC Special Report*.

Hori C, Gaskell J, Igarashi K, Kersten P, Mozuch M, Samejima M, Cullen D. **2014**. Temporal alterations in the secretome of the selective ligninolytic fungus *Ceriporiopsis subvermispora* during

growth on aspen wood reveal this organism's strategy for degrading lignocellulose. *Appl Environ Microbiol.* **80**: 2062–2070.

Ichinose H, Kitaoka T. **2018**. Insight into metabolic diversity of the brown-rot basidiomycete *Postia placenta* responsible for sesquiterpene biosynthesis: semi-comprehensive screening of cytochrome P450 monooxygenase involved in protoilludene metabolism. *Microbial biotechnology* **11**: 952–965.

Justo A, Miettinen O, Floudas D, Ortiz-Santana B, Sjökvist E, Lindner D, Nakasone K, Niemelä T, Larsson KH, Ryvarden L, *et al.* 2017. A revised family-level classification of the Polyporales (Basidiomycota). *Fungal Biology* **121**: 798–824.

Keller NP, Turner G, Bennett JW. **2005**. Fungal secondary metabolism — from biochemistry to genomics. *Nature Reviews Microbiology* **3**: 937–947.

Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND. **2010**. SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genetics and Biology* **47**: 736–741.

Kim K-T, Jeon J, Choi J, Cheong K, Song H, Choi G, Kang S, Lee Y-H. **2016**. Kingdom-Wide Analysis of Fungal Small Secreted Proteins (SSPs) Reveals their Potential Role in Host Association. *Frontiers in Plant Science* **7**: 186.

Kirk P, Cannon P, Minter D, Stalpers J. **2009**. Dictionary of the Fungi. *Mycol Res* **113**: 908–910.

Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canbäck B, Choi C, Cichocki N, Clum A, *et al.* 2015. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nature Genetics* **47**: 410–415.

Kosuta S, Chabaud M, Lougnon G, Gough C, Dénarié J, Barker DG, Bécard G. **2003**. A diffusible factor from arbuscular mycorrhizal fungi induces symbiosis-specific MtENOD11 expression in roots of *Medicago truncatula*. *Plant physiology* **131**: 952–962.

Krah F-S, Bässler C, Heibl C, Soghigian J, Schaefer H, Hibbett DS. **2018**. Evolutionary dynamics of host specialization in wood-decay fungi. *BMC Evolutionary Biology* **18**: 119.

Kumar S. **2010**. Engineering cytochrome P450 biocatalysts for biotechnology, medicine and bioremediation. *Expert Opinion on Drug Metabolism & Toxicology* **6**: 115–131.

Kumar S, Trivedi PK. **2018**. Glutathione S-Transferases: Role in Combating Abiotic Stresses Including Arsenic Detoxification in Plants. *Frontiers in Plant Science* **9**: 751.

Kuo A, Bushnell B, Grigoriev I V. **2014**. Chapter One - Fungal Genomics: Sequencing and Annotation. In: Martin FMBT-A in BR, ed. Fungi. Academic Press, 1–52.

Kuuskeri J, Häkkinen M, Laine P, Smolander O-P, Tamene F, Miettinen S, Nousiainen P, Kemell M, Auvinen P, Lundell T. **2016**. Time-scale dynamics of proteome and transcriptome of the white-rot fungus *Phlebia radiata*: growth on spruce wood and decay effect on lignocellulose. *Biotechnology for Biofuels* **9**: 192.

Lange L, Barrett K, Pilgaard B, Gleason F, Tsang A. **2019**. Enzymes of early-diverging, zoosporic fungi. *Applied Microbiology and Biotechnology* **103**: 6885–6902.

Laurent B, Moinard M, Spataro C, Ponts N, Barreau C, Foulongne-Oriol M. **2017**. Landscape of genomic diversity and host adaptation in *Fusarium graminearum*. *BMC Genomics* **18**: 203.

Lehtiö J, Sugiyama J, Gustavsson M, Fransson L, Linder M, Teeri TT. **2003**. The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proceedings of the National Academy of Sciences* **100**: 484 LP – 489.

Lladó S, López-Mondéjar R, Baldrian P. **2017**. Forest Soil Bacteria: Diversity, Involvement in

Ecosystem Processes, and Response to Global Change. *Microbiology and Molecular Biology Reviews* **81**: e00063-16.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. **2014**. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**: D490-5.

Lundell T, Mäkelä M, Vries RP, Hildén K. **2014b**. Genomics, Lifestyles and Future Prospects of Wood-Decay and Litter-Decomposing Basidiomycota. *Advances in Botanical Research*. 70: 329–370.

Lynd LR, Liang X, Biddy MJ, Allee A, Cai H, Foust T, Himmel ME, Laser MS, Wang M, Wyman CE. **2017**. Cellulosic ethanol: status and innovation. *Current Opinion in Biotechnology* **45**: 202–211.

Mäkelä M.R., Tuomela M., Hatakka A., Hildén K. **2020**. Fungal Laccases and Their Potential in Bioremediation Applications. In: Schlosser D. (eds) Laccases in Bioremediation and Waste Valorisation. *Microbiology Monographs*.

MacDonald J, Goacher RE, Abou-Zaid M, Master ER. **2016**. Comparative analysis of lignin peroxidase and manganese peroxidase activity on coniferous and deciduous wood using ToF-SIMS. *Applied Microbiology and Biotechnology* **100**: 8013–8020.

MacDonald J, Master ER. **2012**. Time-Dependent Profiles of Transcripts Encoding Lignocellulose-Modifying Enzymes of the White Rot Fungus *Phanerochaete carnosa* grown on multiple wood substrates. *Applied and Environmental Microbiology* **78**: 1596–1600.

Mäkipää R, Rajala T, Schigel D, Rinne KT, Pennanen T, Abrego N, Ovaskainen O. **2017**. Interactions between soil- and dead wood-inhabiting fungal communities during the decay of Norway spruce logs. *ISME J* **11**: 1964.

Martínez F, Calleja G, Melero JA, Molina R. **2005**. Heterogeneous photo-Fenton degradation of phenolic aqueous solutions over iron-containing SBA-15 catalyst. *Applied Catalysis B: Environmental* **60**: 181–190.

Martínez de Arano I, Topi C, Pettenella D, Secco L, Masiero M, Follesa M, Fragiacomo M, Carnus J, Lefèvre F, Rigolot E, *et al.* 2018. *A forest-based circular bioeconomy for southern Europe: visions, opportunities and challenges Reflections on the bioeconomy*.

Mathieu Y, Piumi F, Valli R, Aramburu JC, Ferreira P, Faulds CB, Record E. **2016**. Activities of Secreted Aryl Alcohol Quinone Oxidoreductases from *Pycnoporus cinnabarinus* Provide Insights into Fungal Degradation of Plant Biomass. *Applied and Environmental Microbiology* **82**: 2411– 2423.

Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. **2011**. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research* **39**: W339–W346.

Miyauchi S, Te'o VSJ, Nevalainen KMH, Bergquist PL. **2014**. Simultaneous expression of the bacterial *Dictyoglomus thermophilum* xynB gene under three different *Trichoderma reesei* promoters. *New Biotechnology* **31**: 98–103.

Miyauchi S, Navarro D, Grigoriev I V, Lipzen A, Riley R, Chevret D, Grisel S, Berrin J-G, Henrissat B, Rosso M-N. **2016**. Visual comparative omics of fungi for plant biomass deconstruction. *Front Microbiol.* **7**: 1335.

Miyauchi S, Navarro D, Grisel S, Chevret D, Berrin J-G, Rosso M-N. **2017**. The integrative omics of white-rot fungus *Pycnoporus coccineus* reveals co-regulated CAZymes for orchestrated lignocellulose breakdown. *PLOS ONE* **12**: e0175528–e0175528.

Miyauchi S, Rancon A, Drula E, Hage H, Chaduli D, Favel A, Grisel S, Henrissat B, Herpoël-Gimbert I, Ruiz-Dueñas FJ, *et al.* 2018. Integrative visual omics of the white-rot fungus *Polyporus brumalis* exposes the biotechnological potential of its oxidative enzymes for delignifying raw plant biomass. *Biotechnol Biofuels* **11**: 201.

Miyauchi S, Kiss E, Kuo A, Drula E, Kohler A, Sánchez-García M, Morin E, Andreopoulos B, Barry KW, Bonito G, *et al.* 2020. Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. *Nature Communications* **11**: 5125.

Moralejo F, Cardoza R, Gutiérrez S, Lombraña M, Fierro F, Martin J. **2002**. Silencing of the Aspergillopepsin B (pepB) Gene of *Aspergillus awamori* by Antisense RNA Expression or Protease Removal by Gene Disruption Results in a Large Increase in Thaumatin Production. *Applied and environmental microbiology* **68**: 3550–3559.

Morin E, Kohler A, Baker AR, Foulongne-Oriol M, Lombard V, Nagye LG, Ohm RA, Patyshakuliyeva A, Brun A, Aerts AL, *et al.* 2012. Genome sequence of the button mushroom *Agaricus bisporus* reveals mechanisms governing adaptation to a humic-rich ecological niche. *Proceedings of the National Academy of Sciences* **109**: 17501–17506.

Nagamine S, Liu C, Nishishita J, Kozaki T, Sogahata K, Sato Y, Minami A, Ozaki T, Schmidt-Dannert C, Maruyama J-I, *et al.* 2019. Ascomycete *Aspergillus oryzae* Is an Efficient Expression Host for Production of Basidiomycete Terpenes by Using Genomic DNA Sequences. *Applied and environmental microbiology* **85**.

Nagy LG, Riley R, Tritt A, Adam C, Daum C, Floudas D, Sun H, Yadav JS, Pangilinan J, Larsson K-H, *et al.* 2016. Comparative Genomics of Early-Diverging Mushroom-Forming Fungi Provides Insights into the Origins of Lignocellulose Decay Capabilities. *Molecular Biology and Evolution* **33**: 959–970.

Navarro D, Rosso M-N, Haon M, Olivé C, Bonnin E, Lesage-Meessen L, Chevret D, Coutinho PM, Henrissat B, Berrin J-G. **2014**. Fast solubilization of recalcitrant cellulosic biomass by the basidiomycete fungus *Laetisaria arvalisinvolves* successive secretion of oxidative and hydrolytic enzymes. *Biotechnology for Biofuels* **7**: 143.

Nayak T, Szewczyk E, Oakley CE, Osmani A, Ukil L, Murray SL, Hynes MJ, Osmani SA, Oakley BR. **2006**. A Versatile and Efficient Gene-Targeting System for *Aspergillus nidulans*. *Genetics* **172**: 1557 LP – 1566.

Nosil P. **2012**. Ecological speciation. *Oxford University Press*.

O'Leary J, Hiscox J, Eastwood DC, Savoury M, Langley A, McDowell SW, Rogers HJ, Boddy L, Müller CT. **2019**. The whiff of decay: Linking volatile production and extracellular enzymes to outcomes of fungal interactions at different temperatures. *Fungal Ecology* **39**: 336–348.

OECD. **2002**. *Need for research and development programmes in sustainable chemistry*.

Olson A, Aerts A, Asiegbu F, Belbahri L, Bouzid O, Broberg A, Canbäck B, Coutinho PM, Cullen D, Dalman K, *et al.* 2012. Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen. *The New phytologist* **194**: 1001–1013.

Olsson A-M, Salmén L. **2004**. The association of water to cellulose and hemicellulose in paper examined by FTIR spectroscopy. *Carbohydrate Research* **339**: 813–818.

Ottosson E, Nordén J, Dahlberg A, Edman M, JÃ¶nsson M, Larsson K-H, Olsson J, PenttilÃ¤ R, Stenlid J, Ovaskainen O. **2014**. Species associations during the succession of wood-inhabiting fungal communities. *Fungal Ecol.* **11**: 17–28.

Parfrey LW, Lahr DJG, Knoll AH, Katz LA. **2011**. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences* **108**: 13624–13629.

Park H-S, Jun S-C, Han K-H, Hong S-B, Yu J-H. **2017**. Chapter Three - Diversity, Application, and Synthetic Biology of Industrially Important Aspergillus Fungi. In: Sariaslani S, Gadd GMBT-A in AM, eds. Academic Press, 161–202.

Peintner U, Pöder R, Pümpel T. **1998**. The iceman's fungi. *Mycological Research* **102**: 1153–1162.

Pellegrin C, Morin E, Martin FM, Veneault-Fourrey C. **2015**. Comparative analysis of secretomes from ectomycorrhizal fungi with an emphasis on small-secreted proteins. *Front. Microbiol.* **6**.

Pellerin S, Bamière L, Launay C, Martin R, Schiavo M, Angers D, Augusto L, Balesdent J, Basile-Doelsch I, Bellassen V *et al.* 2019. Stocker du carbone dans les sols français, quel potentiel au regard de l'objectif 4 pour 1000 et à quel coût? Synthèse du rapport d'étude. Agence de l'Environnement et de la Maîtrise de l'Energie.

Perotto S, Daghino S, Martino E. **2018**. Ericoid mycorrhizal fungi and their genomes: another side to the mycorrhizal symbiosis? *New Phytologist* **220**: 1141–1147.

Pinedo C, Wang C-M, Pradier J-M, Dalmais B, Choquer M, Le Pêcheur P, Morgant G, Collado IG, Cane DE, Viaud M. **2008**. Sesquiterpene Synthase from the Botrydial Biosynthetic Gene Cluster of the Phytopathogen *Botrytis cinerea*. *ACS Chemical Biology* **3**: 791–801.

Piumi F, Levasseur A, Navarro D, Zhou S, Mathieu Y, Ropartz D, Ludwig R, Faulds CB, Record E. **2014**. A novel glucose dehydrogenase from the white-rot fungus *Pycnoporus cinnabarinus*: production in *Aspergillus niger* and physicochemical characterization of the recombinant enzyme. *Applied Microbiology and Biotechnology* **98**: 10105–10118.

Plett JM, Daguerre Y, Wittulsky S, Vayssières A, Deveau A, Melton SJ, Kohler A, Morrell-Falvey JL, Brun A, Veneault-Fourrey C, *et al.* 2014. Effector MiSSP7 of the mutualistic fungus Laccaria bicolor stabilizes the Populus JAZ6 protein and represses jasmonic acid (JA) responsive genes. *Proceedings of the National Academy of Sciences*.

Popper ZA, Michel G, Hervé C, Domozych DS, Willats WGT, Tuohy MG, Kloareg B, Stengel DB. **2011**. Evolution and Diversity of Plant Cell Walls: From Algae to Flowering Plants. *Annual Review of Plant Biology* **62**: 567–590.

Priya P, Yadav A, Chand J, Yadav G. **2018**. Terzyme: a tool for identification and analysis of the plant terpenome. *Plant Methods* **14**: 4.

Proctor RH, Hohn TM. **1993**. Aristolochene synthase. Isolation, characterization, and bacterial expression of a sesquiterpenoid biosynthetic gene (Ari1) from *Penicillium roqueforti*. *The Journal of biological chemistry* **268**: 4543–4548.

Proctor RH, Hohn TM, McCormick SP. **1995**. Reduced virulence of *Gibberella zeae* caused by disruption of a trichothecene toxin biosynthetic gene. *Molecular plant-microbe interactions : MPMI* **8**: 593–601.

Rajala T, Peltoniemi M, Pennanen T, Mäkipä R. **2012**. Fungal community dynamics in relation to substrate quality of decaying Norway spruce (*Picea abies* [L.] Karst.) logs in boreal forests. *FEMS Microbiol Ecol.* **81**: 494–505.

Rayner ADM, Boddy L, Dowson CG. **1987**. Temporary parasitism of *Coriolus spp.* by *Lenzites betulina*: A strategy for domain capture in wood decay fungi. *FEMS Microbiology Letters* **45**: 53–58.

Riley R, Salamov AA, Brown DW, Nagy LG, Floudas D, Held BW, Levasseur A, Lombard V, Morin E, Otillar R, *et al.* 2014. Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *Proceedings of the National Academy of Sciences* **111**: 9923 LP – 9928.

Rineau F, Roth D, Shah F, Smits M, Johansson T, Canbäck B, Olsen PB, Persson P, Grell MN, Lindquist E, *et al.* 2012. The ectomycorrhizal fungus *Paxillus involutus* converts organic matter in plant litter using a trimmed brown-rot mechanism involving Fenton chemistry. *Environmental Microbiology* **14**: 1477–1487.

Rineau F, Shah F, Smits MM, Persson P, Johansson T, Carleer R, Troein C, Tunlid A. **2013**. Carbon availability triggers the decomposition of plant litter and assimilation of nitrogen by an ectomycorrhizal fungus. *The ISME Journal* **7**: 2010–2022.

Robin C, Andanson A, Saint-Jean G, Fabreguettes O, Dutech C. **2017**. What was old is new again: thermal adaptation within clonal lineages during range expansion in a fungal pathogen. *Molecular Ecology* **26**: 1952–1963.

Rokas A, Mead ME, Steenwyk JL, Raja HA, Oberlies NH. **2020**. Biosynthetic gene clusters and the evolution of fungal chemodiversity. *Natural product reports* **37**: 868–878.

Rubin, E.M. 2008. Genomics of cellulosic biofuels. *Nature* **454**: 841–845.

Ruiz-Dueñas FJ, Lundell T, Floudas D, Nagy LG, Barrasa JM, Hibbett DS, Martínez AT. **2013**. Lignin-degrading peroxidases in Polyporales: an evolutionary survey based on 10 sequenced genomes. *Mycologia* **105**: 1428–1444.

Ruiz-Dueñas FJ, Barrasa JM, Sánchez-García M, Camarero S, Miyauchi S, Serrano A, Linde D, Babiker R, Drula E, Ayuso-Fernández I, *et al.* 2020. Genomic Analysis Enlightens Agaricales Lifestyle Evolution and Increasing Peroxidase Diversity. *Molecular Biology and Evolution*.

Rytioja J, Hildén K, Yuzon J, Hatakka A, de Vries RP, Mäkelä MR. **2014**. Plant-polysaccharide-degrading enzymes from Basidiomycetes. *Microbiology and molecular biology reviews : MMBR* **78**: 614–649.

Sánchez-García M, Ryberg M, Khan FK, Varga T, Nagy LG, Hibbett DS. **2020**. Fruiting body form, not nutritional mode, is the major driver of diversification in mushroom-forming fungi. *Proceedings of the National Academy of Sciences*: 201922539.

Scheller HV, Ulvskov P. **2010**. Hemicelluloses. *Annual Review of Plant Biology* **61**: 263–289.

Schilling JS, Kaffenberger JT, Held BW, Ortiz R, Blanchette RA. **2020**. Using Wood Rot Phenotypes to Illuminate the "Gray" Among Decomposer Fungi . *Frontiers in Microbiology* **11**: 1288.

Schmidt-Dannert C. **2015**. Biosynthesis of Terpenoid Natural Products in Fungi BT - Biotechnology of Isoprenoids. In: Schrader J, Bohlmann J, eds. Cham: Springer International Publishing, 19–61.

Seidl V, Seiboth B. **2010**. *Trichoderma reesei*: genetic approaches to improving strain efficiency. *Biofuels* **1**: 343–354.

Sipos G, Prasanna AN, Walter MC, O'Connor E, Bálint B, Krizsán K, Kiss B, Hess J, Varga T, Slot J, *et al.* 2017. Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi *Armillaria*. *Nature ecology & evolution* **1**: 1931–1941.

Sobel JM, Chen GF, Watt LR, Schemske DW. **2010**. The biology of speciation. *Evolution* **64**: 295–315.

Sokolski S, Bernier-Cardou M, Piché Y, Bérubé JA. **2007**. Black spruce (*Picea mariana*) foliage hosts numerous and potentially endemic fungal endophytes. *Canadian Journal of Forest Research* **37**: 1737–

1747.

Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL. **2007**. *Climate change 2007: the physical science basis*.

Song J, Cui B-K. **2017**. Phylogeny, divergence time and historical biogeography of *Laetiporus* (Basidiomycota, Polyporales). *BMC evolutionary biology* **17**: 102.

Spatafora JW, Aime MC, Grigoriev I V, Martin F, Stajich JE, Blackwell M. **2017**. The Fungal Tree of Life: From Molecular Systematics to Genome-Scale Phylogenies. *The Fungal Kingdom*: 1–34.

Srivastava N, Rawat R, Singh Oberoi H, Ramteke PW. **2015**. A Review on Fuel Ethanol Production From Lignocellulosic Biomass. *International Journal of Green Energy* **12**: 949–960.

Stichnothe H, Bell G, Jørgensen H, Bari I, Haveren J, Lindorfer J, Kepler J, de Jong E. **2020**. Bio-Based Chemicals, A 2020 Update. *IEA Bioenergy*.

Suzuki H, MacDonald J, Syed K, Salamov A, Hori C, Aerts A, Henrissat B, Wiebenga A, VanKuyk PA, Barry K, *et al.* 2012. Comparative genomics of the white-rot fungi*, Phanerochaete carnosa* and *P. chrysosporium*, to elucidate the genetic basis of the distinct wood types they colonize. *BMC Genomics* **13**: 444.

Tabima JF, Trautman IA, Chang Y, Wang Y, Mondo S, Kuo A, Salamov A, Grigoriev I V, Stajich JE, Spatafora JW. **2020**. Phylogenomic analyses of non-Dikarya fungi supports horizontal gene transfer driving diversification of secondary metabolism in the amphibian gastrointestinal symbiont, *Basidiobolus*. *bioRxiv* **9**:3417–3433.

Takayoshi Higuchi. **1998**. The Discovery of Lignin. *Discoveries in Plant Biology* **II**.

Tayeb J. **2019**. *2G Ethanol: Futurol technology is almost on the market*.

Teotia P, Kumar V, Kumar M, Shrivastava N, Varma A. **2016**. Rhizosphere Microbes: Potassium Solubilization and Crop Productivity – Present and Future Aspects BT - Potassium Solubilizing Microorganisms for Sustainable Agriculture. In: Meena VS, Maurya BR, Verma JP, Meena RS, eds. New Delhi: Springer India, 315–325.

Tijerino A, Cardoza RE, Moraga J, Malmierca MG, Vicente F, Aleu J, Collado IG, Gutiérrez S, Monte E, Hermosa R. **2011**. Overexpression of the trichodiene synthase gene tri5 increases trichodermin production and antimicrobial activity in *Trichoderma brevicompactum*. *Fungal genetics and biology : FG & B* **48**: 285–296.

Trejo-Hernández A, Andrade-Domínguez A, Hernández M, Encarnación S. **2014**. Interspecies competition triggers virulence and mutability in *Candida albicans–Pseudomonas aeruginosa* mixed biofilms. *The ISME Journal* **8**: 1974–1988.

Valentín L, Rajala T, Peltoniemi M, Heinonsalo J, Pennanen T, MÃ¤kipÃ¤Ã¤ R. **2014**. Loss of diversity in wood-inhabiting fungal communities affects decomposition activity in Norway spruce wood. *Front Microbiol.* **5**.

Varga T, Krizsán K, Földi C, Dima B, Sánchez-García M, Sánchez-Ramírez S, Szöllősi GJ, Szarkándi JG, Papp V, Albert L, *et al.* 2019. Megaphylogeny resolves global patterns of mushroom evolution. *Nature ecology & evolution* **3**: 668–678.

Veneault-Fourrey C, Commun C, Kohler A, Morin E, Balestrini R, Plett J, Danchin E, Coutinho P, Wiebenga A, de Vries RP, *et al.* 2014. Genomic and transcriptomic analysis of *Laccaria bicolor* CAZome reveals insights into polysaccharides remodelling during symbiosis establishment. *Fungal Genetics and Biology* **72**: 168–181.

Venugopal P, Junninen K, Linnakoski R, Edman M, Kouki J. **2016**. Climate and wood quality have decayer-specific effects on fungal wood decomposition. *Forest Ecology and Management* **360**: 341–351.

Wang C-M, Hopson R, Lin X, Cane DE. **2009**. Biosynthesis of the sesquiterpene botrydial in *Botrytis cinerea*. Mechanism and stereochemistry of the enzymatic formation of presilphiperfolan-8β-ol. *Journal of the American Chemical Society* **131**: 8360–8361.

Wawrzyn GT, Quin MB, Choudhary S, López-Gallego F, Schmidt-Dannert C. **2012**. Draft genome of *Omphalotus olearius* provides a predictive framework for sesquiterpenoid natural product biosynthesis in Basidiomycota. *Chemistry & biology* **19**: 772–783.

WBA. **2016**. Global biomass potential towards 2035. *World Bioenergy association*.

WBA. **2019**. Bioenergy continues to play a prominent role in global energy mix. *World Bioenergy association*.

Weber T, Kim HU. **2016**. The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. *Synthetic and Systems Biotechnology* **1**: 69–79.

Wirth S, Krause K, Kunert M, Broska S, Paetz C, et al. 2021. Function of sesquiterpenes from *Schizophyllum commune* in interspecific interactions. PLOS ONE **16**: e0245623.

Wu W, Tran W, Taatjes CA, Alonso-Gutierrez J, Lee TS, Gladden JM. **2016**. Rapid Discovery and Functional Characterization of Terpene Synthases from Four Endophytic *Xylariaceae*. *PLOS ONE* **11**: e0146983.

Xiao H, Zhong J-J. **2016**. Production of Useful Terpenoids by Higher-Fungus Cell Factory and Synthetic Biology Approaches. *Trends in Biotechnology* **34**: 242–255.

Yoav S, Salame TM, Feldman D, Levinson D, Ioelovich M, Morag E, Yarden O, Bayer EA, Hadar Y. **2018**. Effects of cre1 modification in the white-rot fungus *Pleurotus ostreatus* PC9: altering substrate preference during biological pretreatment. *Biotechnology for Biofuels* **11**: 212.

Zhang J, Presley GN, Hammel KE, Ryu J-S, Menke JR, Figueroa M, Hu D, Orr G, Schilling JS. **2016**. Localizing gene regulation reveals a staggered wood decay mechanism for the brown rot fungus *Postia placenta*. *Proc Natl Acad Sci USA* **113**: 10968–10973.

Zhang F, Anasontzis GE, Labourel A, Champion C, Haon M, Kemppainen M, Commun C, Deveau A, Pardo A, Veneault-Fourrey C, *et al.* 2018. The ectomycorrhizal basidiomycete *Laccaria bicolor* releases a secreted β-1,4 endoglucanase that plays a key role in symbiosis development. *New Phytologist* **220**: 1309–1321.

Zhang J, Silverstein KAT, Castaño JD, Figueroa M, Schilling JS. **2019**. Gene Regulation Shifts Shed Light on Fungal Adaption in Plant Biomass Decomposers. *mBio* **10**.

Zhang C, Chen X, Orban A, Shukal S, Birk F, Too H-P, Rühl M. **2020**. *Agrocybe aegerita* Serves As a Gateway for Identifying Sesquiterpene Biosynthetic Enzymes in Higher Fungi. *ACS Chemical Biology* **15**: 1268–1277.

Zhou S, Raouche S, Grisel S, Navarro D, Sigoillot J-C, Herpoël-Gimbert I. **2015**. Solid-state fermentation in multi-well plates to assess pretreatment efficiency of rot fungi on lignocellulose biomass. *Microbial biotechnology* **8**: 940–949.

Zimmer A, Lang D, Richardt S, Frank W, Reski R, Rensing SA. **2007**. Dating the early evolution of plants: detection and molecular clock analyses of orthologs. *Molecular Genetics and Genomics* **278**: 393–402.

ANNEXES

ANNEXE 1: Distribution of methionine sulfoxide reductases in fungi and evidence for horizontal gene transfers.

Fungi are living in harsh environments that require constant physiological modulations to survive and adapt to various stresses. Those physiological modifications include posttranslational modifications (PTMs) of proteins (e.g. phosphorylation, glycosylation, or oxidation...) that occur after the synthesis of their polypeptide chain (translation), sometimes leading to changes in their activities. While some PTM can be beneficial to the cell, others can cause severe intracellular damages leading to cell death. This latter phenomenon can be caused by the increase in cell concentration of reactive oxygen species (ROS), which are the singlet oxygen (O_2), the superoxide anion ($O^{2\bullet-}$), the hydrogen peroxide (H_2O_2) and the hydroxyl radical ($HO^{\bullet}$). Those molecules have the ability to oxidize amino acids of proteins, especially sulfur-containing amino acids, e.g cysteine (Cys) and methionine (Met). Fortunately, organisms have adapted specific enzymatic systems to repair damages by reversing the redox modifications affecting Cys and Met. The enzymes that reduce the oxidized methionine (MetO) to Met are called methionine sulfoxide reductases (Msr). Overall, the functions of MSR proteins were well studied in animals, bacteria and to a lesser extent in plants but far less explored in fungi.

In this paper, we performed a global genomic search for Msr genes in almost 700 genomes covering all the fungal kingdom. We show that most of the fungi possess one gene coding for each of the three Msr type: MsrA, MsrB and fRMsr. We also identified a few fungi devoid of any Msr genes at all. Sequence characterization and phylogenetic analyses revealed that the Msrs are very well conserved throughout the fungal kingdom and allowed also to identify unusual sequences with potentially interesting characteristics and uncovered bacteria to fungi horizontal gene transfers.

This work is presented in the manuscript “Hage H, Rosso MN, Tarrago L. Characterization of methionine sulfoxide reductases families in fungi reveals the conservation of free-methionine-R-sulfoxide reductase (fRMsr) in multicellular eucaryotes.” (*In preparation*).

Supplementary data are available at the following link:
https://figshare.com/s/7664ac6d335c1ac8a5fa

Distribution of methionine sulfoxide reductases in fungi and evidence for horizontal gene transfers.

Hayat Hage[1], Marie-Noëlle Rosso[1], Lionel Tarrago[1,*]

[1]Biodiversité et Biotechnologie Fongiques, UMR1163, INRAE, Aix Marseille Université, Marseille, France.

*Correspondence: Lionel Tarrago (lionel.tarrago@inrae.fr)

Running title: Methionine sulfoxide reductases in fungi

Keywords: comparative genomics, fungi, genomes, horizontal gene transfers, methionine sulfoxide, methionine sulfoxide reductase, protein oxidation, thiol oxidoreductases.

Abstract

Methionine, either as free amino acid or included in proteins, can be oxidized into methionine sulfoxide (MetO), which exists as *R* and *S* diastereomers. Almost all characterized organisms possess methionine sulfoxide reductase (Msr) enzymes to reduce MetO back to Met. The thiol-oxidoreductases MsrA and MsrB reduce the *S* and *R* diastereomers of MetO with strict stereospecificity and are found in almost all organisms. Another thiol-oxidoreductase, the free-Methionine-*R*-sulfoxide reductase (fRMsr) reduces only the *R* MetO diastereomer of the free amino acid and was identified so far in prokaryotes and a few unicellular eukaryotes. Moreover, some bacteria possess molybdenum-containing enzymes reducing the MetO, either free or included in proteins. All these Msrs play key roles in the protection of organisms against oxidative stress. Fungi are heterotrophic unicellular and multicellular eukaryotes colonizing all niches on Earth and playing fundamental functions in organic matter recycling or as symbionts or pathogens of numerous organisms. However, our knowledge on fungal Msrs is scarce. Here, we performed a global genomic search for *msr* genes in almost 700 genomes covering all the fungal kingdom. We show that most of the fungi possess one gene coding for each type of thiol-oxidoreductase, MsrA, MsrB, and fRMsr. This demonstrates that the fRMsr is present in multicellular eukaryotes. We also identified several fungi devoid of any *msr* genes. Sequence inspection and phylogenetic analyses allowed us to identify unusual sequences with potentially interesting characteristics and to uncover horizontal gene transfers from bacteria to fungi.

Introduction

Life in presence of dioxygen (O_2) necessarily exposes the biological systems to oxidant molecules. Due to its high reactivity, O_2 can be converted in *reactive oxygen species* (ROS), like hydrogen peroxide (H_2O_2), which plays key roles through redox modifications of macromolecules in physiological and pathological contexts (Sies and Jones, 2020). Either as the free amino acid or as a residue included in a protein, Met can be oxidized into Met sulfoxide (MetO) by the addition of an oxygen atom on the sulfur of the lateral chain. MetO exists as diastereomers *R* and *S* (Met-*R*-O and Met-*S*-O, respectively) (Sharov et al., 1999), which can be reduced back to Met by the action of (seleno)thiol oxidoreductases called methionine sulfoxide reductases (Msr). The two main types are the MsrA and MsrB which display strict stereoselectivities toward the *S*- and the *R*-diastereomer of MetO, respectively (Brot et al., 1981; Grimaud et al., 2001; Sharov and Schöneich, 2000). Whereas MsrA reduces MetO as the free amino acid or as a residue in protein with similar catalytic efficiencies, MsrB generally reduces efficiently only the protein-bound MetO (Tarrago et al., 2012; Tarrago and Gladyshev, 2012). Another enzyme, the fRMsr (for *'Free methionine-(*R*)-sulfoxide reductase'*), has been shown to reduce specifically the free form of Met-*R*-O and previous genomics searches reported its presence in bacteria and in the genome of some unicellular eukaryotes (Gruez et al., 2010; Le et al., 2009; Lin et al., 2007). Bacteria can also possess several molybdoenzymes that may reduce exclusively the free-MetO (Dhouib et al., 2016; Ezraty et al., 2005) or both the free and protein-bound MetO (Andrieu et al., 2020; Gennaris et al., 2015; Tarrago et al., 2020, 2018).

Despite the lack of sequence and structure similarities, MsrA, MsrB and fRMsr catalyze the reduction of MetO using a similar 3-steps-mechanism in most cases (Lin et al., 2007; Lowther et al., 2002) : i) a *'catalytic'* Cys (or selenocysteine, Sec) reduces the target MetO and is converted into a sulfenic (or selenic) acid, ii) an internal *'resolving'* Cys reduces it through the formation of an intramolecular disulfide bond, and finally iii) the oxidized Msr is regenerated through disulfide exchange with a thioredoxin (Tarrago and Gladyshev, 2012). Variations of this mechanism exist, with some Msrs devoid of any resolving Cys, for which the sulfenic acid is directly reduced by an external reducer (Kim, 2012; Lee et al., 2015; Tarrago et al., 2009b, 2010). Alternatively, in other Msrs, a disulfide exchange occurs with a second internal resolving Cys before the regeneration by the thioredoxin (Boschi-Muller et al., 2000; Gruez et al., 2010; Lowther et al., 2000; Rouhier et al., 2007; Tarrago et al., 2009b, 2012).

Genome mining analyses indicated that *msrA* and *msrB* genes originated from bacteria and are now present in all organisms across the tree of life, with few exceptions (Delaye et al., 2007; Tarrago et al., 2009a; Zhang and Weissbach, 2008). Most organisms have few *msr* genes, generally one for each type (Delaye et al., 2007; Tarrago et al., 2009a; Zhang and Weissbach, 2008), but this number can increase up to 5 *msrA* and 9 *msrB* genes in the plant *Arabidopsis thaliana* (Tarrago et al., 2009a). Moreover, some bacteria encode a bifunctional MsrA/MsrB fusion and some others lack a MsrB, but no organism of any kind was so far described with only

MsrB (Delaye et al., 2007; Zhang and Weissbach, 2008). Finally, only very few organisms do not possess any *msr* gene at all, such as a few endosymbiotic or obligatory parasitic bacteria and some archaea (Delaye et al., 2007; Zhang and Weissbach, 2008). Very interestingly, the only eukaryote for which the absence of *msr* was described is the fungus *Encephalitozoon cuniculi*, a microsporidium having a remarkably reduced genome (∼2.9 Mb) and living as intracellular parasite of mammals (Delaye et al., 2007; Katinka et al., 2001).

The conservation of typical MsrAs and MsrBs in almost all known organisms argues for the general and critical role of MetO reduction in cellular metabolism, and numerous studies showed that Msrs are involved in the protection against oxidative stress and the regulation of protein functions. Schematically, Msr roles in the protection against oxidative injuries occur through two main functions: i) the repair of oxidized proteins, and ii) an antioxidant function through the scavenging of ROS by cyclic Met oxidation and reduction. Moreover, the reversible Met oxidation was shown to act as a regulating post-translational modification implicated in the activation of enzymes and transcription factors and in the regulation of protein-protein interactions (Drazic et al., 2013; Erickson et al., 2008; Jiang et al., 2020; Lee et al., 2013). The role of fRMsr has been far less studied, but it is suspected to have an antioxidant function by reducing free MetO and maintaining the pool of Met available for protein synthesis and for the production of sulfur-containing metabolites (Le et al., 2009; Lee et al., 2018). Overall, these aspects were well established in animals, bacteria and plants, and have recently been reviewed (Jiang and Moskovitz, 2018; Lim et al., 2018; Lourenço dos Santos et al., 2018; Rey and Tarrago, 2018; Singh et al., 2018).

Fungi are heterotrophic eukaryotes that colonize virtually all niches on Earth and exist as unicellular or multicellular organisms. They can have numerous lifestyles, either as free living organisms playing key roles in organic and inorganic matter cycling, or as symbionts or pathogens with crucial impacts on plant and animal health (Naranjo-Ortiz and Gabaldón, 2019). Moreover, fungi are invaluable sources for applications in food and drug industries (Meyer et al., 2020). However, despite the fundamental roles of fungi in life, the effects of Met oxidation and the functions of Msrs was largely overlooked in these eukaryotes. The Msr system was mainly characterized in the yeast *Saccharomyces cerevisiae*, which possesses one Msr of each type (Kryukov et al., 2002; Le et al., 2009; Tarrago et al., 2012). MsrA is located in the cytosol, MsrB is found in mitochondria and in the cytosol and fRMsr was localized both in the cytosol and the nucleus (Kaya et al., 2010; Nicklow and Sevier, 2020). Genetic manipulations have shown that these Msrs play a role in protection against oxidative stress and in maintaining the yeast's lifespan (Koc et al., 2004; Kryukov et al., 2002; Le et al., 2009; Moskovitz et al., 1998; Tarrago et al., 2012). Consistently, the overexpression of *msrA* in the basidiomycete *Pleurotus ostreatus* and of *msrB* in the yeast *Schizosaccharomyces pombe* improved the resistance to oxidative constraints (Jo et al., 2014; Yin et al., 2015). Moreover, the filamentous fungus *Aspergillus nidulans* possesses one *msrA* and one *msrB* genes, and the deletion of one or both

genes increases the sensitivity of the fungus to oxidative treatments (Soriani et al., 2009). Few proteins with oxidized Met residues have been characterized in fungi, but remarkably interesting findings have been obtained. For instance, in *S. cerevisiae* it has been shown that the reversible oxidation of Met regulates the oligomerization state of the ataxin-2 protein and the activity of the co-chaperone Fes1 (Kato et al., 2019; Nicklow and Sevier, 2020). Moreover, Met oxidation enhanced the activity of an α-galactosidase in *Trichoderma reesei* (Kachurin et al., 1995). Finally, in *A. nidulans*, it was demonstrated that the nuclear localization of the nitrate-responsive transcription factor NirA was actively regulated through cyclic Met oxidation (Gallmetzer et al., 2015). These data indicate that, as for other organisms, fungal Msrs must play key roles under many conditions involving oxidative stress such as biotic interactions or abiotic constraints (Breitenbach et al., 2015).

In this study, we searched for *msr* genes in about 700 published fungal genomes. We show that the great majority of fungi have one *msrA* and one *msrB* genes. Moreover, we identified fRMsr genes in almost all the fungi analyzed and thereby demonstrate that the enzyme is present in multicellular eukaryotes. Finally, using a phylogenetic analysis of fungal Msrs and a close inspection of sequence features, we identified Msrs with unusual sequence characteristics and uncovered horizontal gene transfers from bacteria to fungi.

Results

Search for Msrs in the fungal kingdom

Currently, around 136,000 species of fungi are known and classified into 9 phyla (Figure 1) (James et al., 2020; Naranjo-Ortiz and Gabaldón, 2019). Together, the Ascomycota and the Basidiomycota form the subkingdom Dikarya, which regroups more than 97 % of all described fungal species (~84,000 species and ~48,000 species, respectively) (James et al., 2020). Both phyla contain three monophyletic subphyla (Pucciniomycotina, Ustilaginomycotina and Agaricomycotina for the Basidiomycota, and Pezizomycotina, Saccharomycotina and Taphrinomycotina for the Ascomycota). The other seven phyla are described as *'early-diverging fungi'* from which the Glomeromycota and Mucoromycota are the most closely related to dikarya, and the Opisthosporidia is the most basal branch of Fungi (James et al., 2020) (Figure 1).

Using the *S. cerevisiae* MsrA, MsrB and fRMsr protein sequences as queries, we used the BLASTP and TBLASTN software suites to search for *msr* genes in 682 available genomes in the MycoCosm database (Grigoriev et al., 2014). The selected genomes were from 595 species that spanned the kingdom Fungi (Figure 1), with 65% Ascomycota species, 25% Basidiomycota species and 10 % early-diverging fungi. We found that the very great majority of these genomes

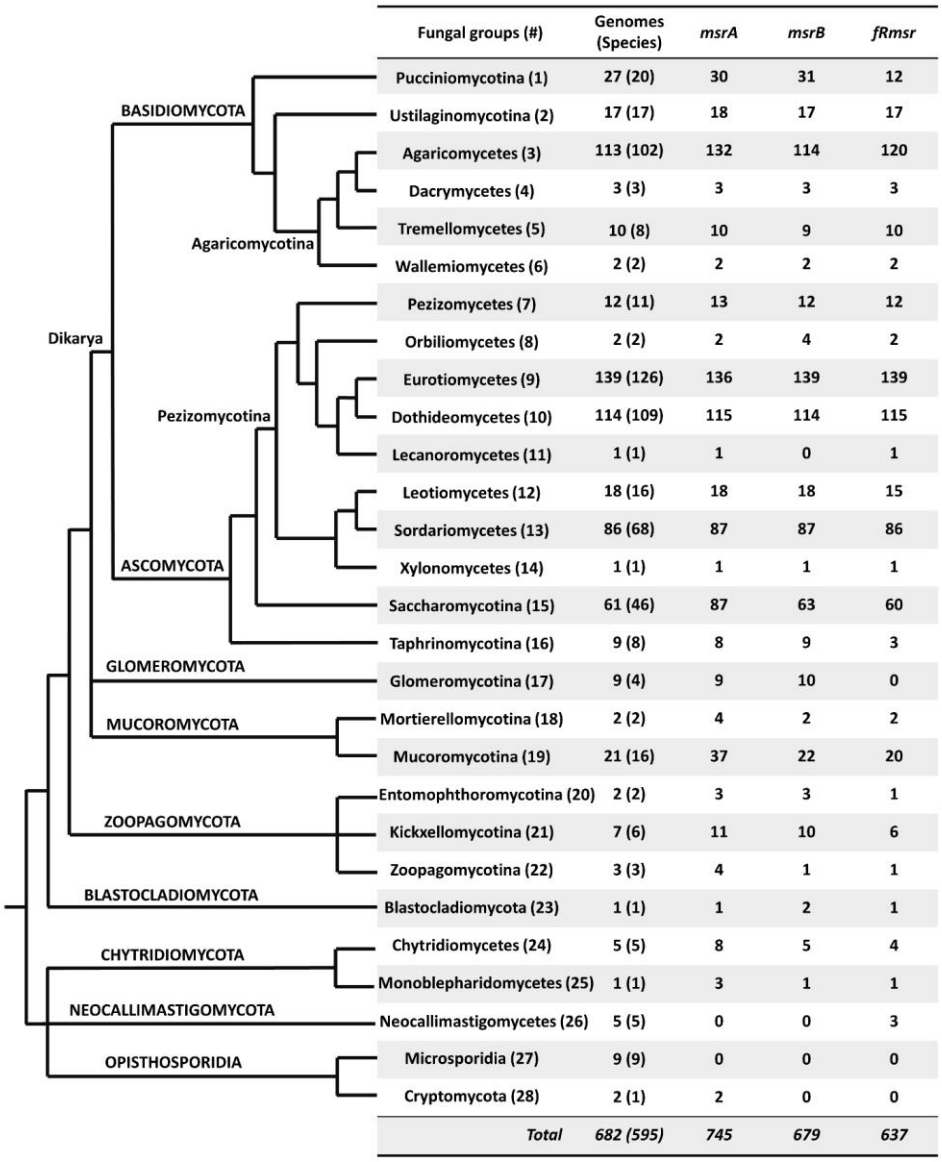
contained one gene coding for a MsrA and one coding for a MsrB (Figure 1), indicating that most fungi have a simple Msr system dedicated to protein oxidation repair, similarly to most other known organisms (Delaye et al., 2007; Zhang and Weissbach, 2008). Very interestingly, we found that a *fRmsr* gene was present in almost all analyzed genomes (Figure 1). The distribution of *fRmsr* across the fungal kingdom clearly showed that the presence of fRMsr is not limited to bacteria and unicellular eukaryotes as previously described (Gruez et al., 2010; Le et al., 2009; Lin et al., 2007). Of note, the search for homologs of the bacterial molybdoenzymes able to reduce MetO gave no significant hits (data not shown).

We found only 60 genomes, corresponding to 50 species, in which one or more Msr types were absent (Table 1). Among them, the nine Microsporidia species, two Neocallimastigomycetes species (out of five species analyzed) and the Taphrinomycotina species *Pneumocystis jirovecii* (out of eight species analyzed) lacked all three Msr types. These species are either obligate intracellular parasites (Microsporidia and *Pneumocystis jirovecii*) or live in the anaerobic conditions of the animal gut (*Piromyces finnis* and *Piromyces* sp. E2) (Naranjo-Ortiz and Gabaldón, 2019; Skalski et al., 2015). Such lifestyles, which may preclude ROS formation and protein oxidation, could have allowed the loss of the *msr* genes by the lack of selection pressure. Interestingly, the three other Neocallimastigomycetes species (*Anaeromyces robustus, Neocallimastix californiae* G1 and *Orpinomyces* sp.), also living in anaerobic conditions (James et al., 2020), do not possess MsrA nor MsrB but have a gene coding for a fRMsr (Table 1). These results show that, additionally to *Encephalitozoon cuniculi* previously identified as the unique eukaryote lacking both MsrA and MsrB (Delaye et al., 2007), 12 other eukaryote species are devoid of any protein-bound MetO reduction enzyme (Table 1).

Also, we found six species (six genomes) lacking the MsrA but having a MsrB and a fRMsr (Table 1). These are the agaricomycetes *Moniliophthora perniciosa*, the Eurotiomycetes *Cladophialophora immunda*, *Penicillium coprophilum* and *Penicillium flavigenum*, the Sordiariomycetes *Magnaporthiopsis poae* and the Saccharomycotina *Saturnispora dispora*. To our knowledge, they constitute the first species from all kingdoms described to have only a MsrB to reduce and repair oxidized proteins as none have been found in previous global genomic characterizations (Delaye et al., 2007; Zhang and Weissbach, 2008). Of note, we did not find fungal species having only a MsrB, as all these six species possessed a fRMsr too (Table 1). We found that the five species *Trichosporon asahii* (Tremellomycetes), *Clohesyomyces aquaticus* (Dothideomycetes), *Cladonia grayi* (Lecaronomycetes) and *Piptocephalis cylindrospora* (Zoopagomycotina) possessed a MsrA and a fRMsr but lacked a MsrB and that the Zoopagomycotina *Syncephalis pseudoplumigaleata* had a MsrA only (Table 1). Finally, we found only 27 species (corresponding to 37 genomes) having both a MsrA and a MsrB but lacking a fRMsr (Table 1). These species were sporadically dispersed among the different fungal groups, but three groups stood out as remarkable: the Glomeromycotina for which none of the four species analyzed had a fRMsr, the Taphrinomycotina for which six species out of eight analyzed

(including all the *Schizosaccharomyces* species analyzed) were devoid of fRMsr, and the Pucciniomycotina for which all the *Melampsora* and *Puccinia* species analyzed here lacked a fRMsr (Table 1).

Altogether, these results showed that the great majority of fungi possess one gene coding for each protein-repairing MsrA and MsrB, as well as one fRMsr gene.



Fungal groups (#)	Genomes (Species)	*msrA*	*msrB*	*fRmsr*
Pucciniomycotina (1)	27 (20)	30	31	12
Ustilaginomycotina (2)	17 (17)	18	17	17
Agaricomycetes (3)	113 (102)	132	114	120
Dacrymycetes (4)	3 (3)	3	3	3
Tremellomycetes (5)	10 (8)	10	9	10
Wallemiomycetes (6)	2 (2)	2	2	2
Pezizomycetes (7)	12 (11)	13	12	12
Orbiliomycetes (8)	2 (2)	2	4	2
Eurotiomycetes (9)	139 (126)	136	139	139
Dothideomycetes (10)	114 (109)	115	114	115
Lecanoromycetes (11)	1 (1)	1	0	1
Leotiomycetes (12)	18 (16)	18	18	15
Sordariomycetes (13)	86 (68)	87	87	86
Xylonomycetes (14)	1 (1)	1	1	1
Saccharomycotina (15)	61 (46)	87	63	60
Taphrinomycotina (16)	9 (8)	8	9	3
Glomeromycotina (17)	9 (4)	9	10	0
Mortierellomycotina (18)	2 (2)	4	2	2
Mucoromycotina (19)	21 (16)	37	22	20
Entomophthoromycotina (20)	2 (2)	3	3	1
Kickxellomycotina (21)	7 (6)	11	10	6
Zoopagomycotina (22)	3 (3)	4	1	1
Blastocladiomycota (23)	1 (1)	1	2	1
Chytridiomycetes (24)	5 (5)	8	5	4
Monoblepharidomycetes (25)	1 (1)	3	1	1
Neocallimastigomycetes (26)	5 (5)	0	0	3
Microsporidia (27)	9 (9)	0	0	0
Cryptomycota (28)	2 (1)	2	0	0
Total	*682 (595)*	*745*	*679*	*637*

Figure 1 Fungal phylogenetic tree and counts of msrA, msrB and fRmsr genes identified in fungal genomes. The simplified phylogenetic tree was built according to (James et al., 2020; Naranjo-Ortiz and Gabaldón, 2019). The fungal groups are representatives of the genomes available in the MycoCosm database (Grigoriev et al., 2014) (https://mycocosm.jgi.doe.gov/mycocosm/home).

Table 1 Fungal genomes lacking one or more Msr types.

Genome	MsrA	MsrB	fRMsr	Lifestyle	Nb. Genes (Genome size in Mbp)	Reference
Microsporidia (Group 27)						
Antonospora locustae HM-2013	N	N	N	Intracellular parasite of metazoans	2,606 (6)	(Corradi et al., 2007)
Encephalitozoon cuniculi GB-M1	N	N	N	Intracellular parasite of metazoans	1,996 (2)	(Peyretaillade et al., 2009)
Encephalitozoon hellem ATCC 50504	N	N	N	Intracellular parasite of metazoans	1,847 (2)	(Pombert et al., 2012)
Encephalitozoon intestinalis ATCC 50506	N	N	N	Intracellular parasite of metazoans	1,833 (2)	(Corradi et al., 2010)
Encephalitozoon romaleae SJ-2008	N	N	N	Intracellular parasite of metazoans	1,831 (2)	(Pombert et al., 2012)
Enterocytozoon bieneusi H348	N	N	N	Intracellular parasite of metazoans	3,632 (4)	(Akiyoshi et al., 2009)
Mitosporidium daphniae UGP3	N	N	N	Intracellular parasite of metazoans	3,330 (6)	(Haag et al., 2014)
Nematocida parisii ERTm1	N	N	N	Intracellular parasite of metazoans	2,661 (4)	(Cuomo et al., 2012)
Nosema ceranae BRL01	N	N	N	Intracellular parasite of metazoans	2,060 (8)	(Cornman et al., 2009)
Neocallimastigomycetes (Group 26)						
Piromyces finnis v3.0	N	N	N	anaerobic (gut)	10,992 (56)	(Haitjema et al., 2017)
Piromyces sp. E2 v1.0	N	N	N	anaerobic (gut)	14,648 (71)	(Haitjema et al., 2017)
Anaeromyces robustus v1.0	N	N	Y	anaerobic (gut)	12,832 (72)	(Haitjema et al., 2017)
Neocallimastix californiae G1 v1.0	N	N	Y	anaerobic (gut)	20,219 (193)	(Haitjema et al., 2017)
Orpinomyces sp.	N	N	Y	anaerobic (gut)	18,936 (101)	(Youssef et al., 2013)
Taphrinomycotina (Group 16)						
Pneumocystis jirovecii	N	N	N	obligate intracellular parasite	3,520 (8)	(Cissé et al., 2012)
Neolecta irregularis DAH-1 v1.0	Y	Y	N	aerobic	6,674 (15)	(Nguyen et al., 2017)
Schizosaccharomyces cryophilus OY26	Y	Y	N	aerobic	5,180 (12)	(Rhind et al., 2011)
Schizosaccharomyces japonicus yFS275	Y	Y	N	aerobic	4,878 (12)	(Rhind et al., 2011)
Schizosaccharomyces octosporus yFS286	Y	Y	N	aerobic	4,986 (12)	(Rhind et al., 2011)
Schizosaccharomyces pombe	Y	Y	N	aerobic	5,134 (13)	(Wood et al., 2002)
Agaricomycetes (Group 3)						
Moniliophthora perniciosa FA553	N	Y	Y	aerobic (plant pathogen)	13,560 (18)	(Mondego et al., 2008)
Eurotiomycetes (Group 9)						
Penicillium coprophilum IBT 31321	N	Y	Y		8,999 (28)	(Nielsen et al., 2017)
Penicillium flavigenum IBT 14082	N	Y	Y		10,994 (33)	(Nielsen et al., 2017)
Penicillium polonicum IBT 4502	N	Y	Y		10,694 (32)	(Nielsen et al., 2017)
Aspergillus bombycis NRRL 26010	Y	Y	N		12,265 (37)	(Moore et al., 2016)
Aspergillus zonatus v1.0	Y	Y	N		9,886 (29)	(de Vries et al., 2017)
Sordariomycetes (Group 13)						

Magnaporthiopsis poae ATCC 64411	N	Y	Y	plant pathogen	12,335 (40)	(Okagaki et al., 2015)
Pochonia chlamydosporia 170	Y	Y	N	animal pathogen	14,204 (44)	(G. Wang et al., 2016)
Saccharomycotina (Group 15)						
Saturnispora dispora NRRL Y-1447	N	Y	Y	aerobic	4,881 (10)	(Krassowski et al., 2018)
Tremellomycetes (Group 5)						
Trichosporon asahii var. asahii CBS 2479	Y	N	Y	human pathogen	8,300 (24)	(Yang et al., 2003)
Trichosporon asahii var. asahii CBS 8904	Y	N	Y	human pathogen	8,507 (25)	(Yang et al., 2012)
Dothideomycetes (Group 10)						
Clohesyomyces aquaticus v1.0	Y	N	Y	aerobic	15,810 (50)	(Mondo et al., 2017)
Lecaronomycetes (Group 11)						
Cladonia grayi Cgr/DA2myc/ss v2.0	Y	N	Y	lichen	11,389 (35)	(Armaleo et al., 2019)
Zoopagomycotina (Group 20)						
Piptocephalis cylindrospora RSA 2659 single-cell v3.0	Y	N	Y	aerobic	4,301 (11)	(Ahrendt et al., 2018)
Syncephalis pseudoplumigaleata Benny S71-1 single-cell v1.0	Y	N	N	mycoparasite	6,123 (16)	(Ahrendt et al., 2018)
Thamnocephalis sphaerospora RSA 1356 single-cell v1.0	Y	Y	N	mycoparasite	6,857 (18)	(Ahrendt et al., 2018)
Pucciniomycotina (Group 1)						
Melampsora allii-populina 12AY07 v1.0	Y	Y	N	plant pathogen	23,089 (336)	Unpublished
Melampsora americana R15-033-03 v1.0	Y	Y	N	plant pathogen	15,984 (112)	Unpublished
Melampsora larici-populina v2.0	Y	Y	N	plant pathogen		(Duplessis et al., 2011)
Melampsora lini CH5	Y	Y	N	plant pathogen	16,335 (190)	(Nemri et al., 2014)
Puccinia coronata avenae 12NC29	Y	Y	N	plant pathogen	28,270 (166)	(Nazareno et al., 2018)
Puccinia coronata avenae 12SD80	Y	Y	N	plant pathogen	26,323 (150)	(Nazareno et al., 2018)
Puccinia graminis f. sp. tritici v2.0	Y	Y	N	plant pathogen	15,979 (89)	(Duplessis et al., 2011)
Puccinia striiformis f. sp. tritici 104 E137 A-	Y	Y	N	plant pathogen	30,249 (157)	(Schwessinger et al., 2018)
Puccinia striiformis f. sp. tritici PST-130	Y	Y	N	plant pathogen	18,021 (65)	(Cantu et al., 2011)
Puccinia striiformis f. sp. tritici PST-78 v1.0	Y	Y	N	plant pathogen	20,482 (117)	(Cuomo et al., 2017)
Puccinia triticina 1-1 BBBD Race 1	Y	Y	N	plant pathogen	15,685 (135)	(Cuomo et al., 2017)
Leotiomycetes (Group 12)						
Blumeria graminis f. sp. hordei DH14	Y	Y	N	plant pathogen	7,118 (125)	(Frantzeskakis et al., 2018)
Blumeria graminis f. sp. hordei Race1	Y	Y	N	plant pathogen	7,239 (116)	(Frantzeskakis et al., 2018)
Blumeria graminis f. sp. tritici 96224	Y	Y	N	plant pathogen	6,525 (159)	(Wicker et al., 2013)
Glomeromycotina (Group 17)						
Gigaspora rosea v1.0	Y	Y	N	Intracellular? (plant symbiont)	31,291 (568)	(Morin et al., 2019)
Rhizophagus cerebriforme DAOM 227022 v1.0	Y	Y	N	Intracellular? (plant symbiont)	21,549 (137)	(Morin et al., 2019)
Rhizophagus diaphanus v1.0	Y	Y	N	Intracellular? (plant symbiont)	23,252(129)	(Morin et al., 2019)
Rhizophagus irregularis A1 v1.0	Y	Y	N	Intracellular? (plant symbiont)	26,659 (126)	(Chen et al., 2018)
Rhizophagus irregularis A4 v1.0	Y	Y	N	Intracellular? (plant symbiont)	25,760 (138)	(Chen et al., 2018)

Rhizophagus irregularis A5 v1.0	Y	Y	N	Intracellular? (plant symbiont)	26,585 (131)	(Chen et al., 2018)
Rhizophagus irregularis B3 v1.0	Y	Y	N	Intracellular? (plant symbiont)	25,164 (125)	(Chen et al., 2018)
Rhizophagus irregularis C2 v1.0	Y	Y	N	Intracellular? (plant symbiont)	26,756 (123)	(Chen et al., 2018)
Rhizophagus irregularis DAOM 197198 v2.0	Y	Y	N	Intracellular? (plant symbiont)	26,183 (137)	(Chen et al., 2018)
Mucoromycotina (Group 19)						
Jimgerdemannia lactiflua OSC166217	Y	Y	N	Ectomycorhize	12,651 (180)	(Chang et al., 2019)
Lichtheimia corymbifera JMRC:FSU:9682	Y	Y	N	human pathogen	13,404 (34)	(Schwartze et al., 2014)
Entomophthoromycotina (Group 21)						
Basidiobolus meristosporus CBS 931.73 v1.0	Y	Y	N	aerobic	16,111 (89)	(Mondo et al., 2017)
Kickxellomycotina (Group 22)						
Dimargaris cristalligena RSA 468 single-cell v1.0	Y	Y	N	mycoparasite	7,456 (31)	(Ahrendt et al., 2018)
Chytridiomycetes (Group 24)						
Caulochytrium protostelioides ATCC 52028 v1.0	Y	Y	N	mycoparasite	6,168 (22)	(Ahrendt et al., 2018)

Characterization of fungal Msrs protein sequences

In several organisms, the catalytic residue of a MsrA is located in a G[C/U]FW motif, located in the N-terminal part of the protein (Antoine et al., 2006; Boschi-Muller et al., 2000; Kim et al., 2006; Lowther et al., 2002; Ma et al., 2011; Tarrago et al., 2012). The great majority of known MsrAs possess a catalytic Cys, with a few Sec-containing proteins mostly found in some insects, marine organisms or unicellular algae (Kim et al., 2006; Tarrago et al., 2009a). The resolving Cys of a MsrA is located in the C-terminal part of the protein, but the number of Cys involved in the regeneration mechanisms and their precise positions vary (Boschi-Muller et al., 2000; Lowther et al., 2000; Ma et al., 2011; Rouhier et al., 2007; Tarrago et al., 2012). To determine whether the fungal MsrAs share these properties, we analyzed the 709 full length MsrAs we identified in the fungal genomes. The fungal MsrAs ranged from 142 to 322 amino acids, and all of them had a single MsrA domain (Supplementary data 1A). The length differences were mostly due to the presence of N-terminal extensions of variable sizes, indicating the possible presence of secretion signal peptides for protein distribution in subcellular compartments (Supplementary data 1A). We used several targeting predictions programs (see *Methods* section) to evaluate the potential subcellular localization of the fungal MsrAs (Supplementary data 1A). Most MsrAs were predicted to be localized into the cytoplasm (45 %) or in the mitochondria (44 %) (Figure 2A & Supplementary data 1A). The other sequences were predicted to be secreted (6%), to be localized in other compartments or not clearly assigned to any localization (6%) (Figure 2A & Supplementary data 1A). The alignment of the primary sequence of the fungal MsrAs revealed that the great majority (~ 90 %) possessed common features with the previously characterized

S. cerevisiae MsrA (Ma et al., 2011; Tarrago et al., 2012) (Figure 3A & Supplementary data 1B). The catalytic Cys (position 25 in the *S. cerevisiae* MsrA) was located in the conserved motif ^{24}GCFW^{27}. The Tyr^{64}, Glu^{76}; Asp^{111} and Tyr^{116} residues, involved in substrate stabilization and catalysis together with the Phe^{26} and Trp^{27} in *Neisseria meningitidis* MsrA (Antoine et al., 2006), were also conserved. The residues Gly^{47}, His^{100}, Gln^{108}, Gly^{113}, His^{163} and Tyr^{166} were also strictly conserved. Finally, the Cys^{176}, previously identified as resolving Cys (Ma et al., 2011; Tarrago et al., 2012), was included in the ^{173}GYXC^{176} motif (Figure 3A & Supplementary data 1B). Because of their predominance in all the fungal kingdom, we defined the fungal MsrAs having these properties as *'canonical'* sequences. The fungal MsrAs that didn't match these sequence properties were defined as *'non-canonical'* MsrAs (Supplementary data 1B, C). Particularly, we observed that in ~ 5 % of the identified sequences, the Phe^{26} residue in the ^{24}GCFW^{27} motif containing the catalytic Cys was substituted by a Tyr. We also observed the replacement of Asp^{111} by an Asn residue in few sequences (~ 2 %) (Supplementary data 1B). Because of the similarities between the residues, these two last substitutions should likely not prevent the catalysis. Moreover, some variations were also observed for the resolving Cys (Supplementary data 1C). Fourteen sequences (~ 2 %) lacked the conserved ^{173}GYXC^{176} motif but possessed 2 to 4 Cys in a Q[C/S/K]X_2KX[C/N][C/X]XI[R/L]CYG motif similar to those containing the Cys involved in the regeneration process of poplar MsrAs (Rouhier et al., 2007). Some other sequences possessed a Cys residue in the C-terminal region, in non-conserved motif or lacked any potential resolving Cys (Supplementary data 1C). Finally, a special case could be made for MsrAs from the early-diverging fungus *Gonapodya prolifera* (Monoblepharidomycetes). We identified in this genome three non-canonical MsrAs, two of which with the catalytic Cys substituted by a Sec (Mariotti et al., 2019). Each had another Cys, but not in the conserved position of the resolving Cys in canonical fungal MsrAs. These two MsrA sequences had high similarity with the Sec-MsrAs from *Alkaliphilus oremlandii* and *Chlamydomonas reinhardtii*, previously shown to use only the Sec and no other Cys for the regeneration of its activity (Kim et al., 2006; Lee et al., 2015) (Supplementary data 1D).

In the case of MsrBs, previous biochemical characterizations demonstrated that the catalytic Cys is located in a RXCXN motif in the C-terminal part of the protein (Kim and Gladyshev, 2005; Lowther et al., 2002; Neiers et al., 2007; Olry et al., 2002; Ranaivoson et al., 2009; Tarrago et al., 2009b). Mammals express a Sec-containing form in which the Asp is replaced by a Phe (Kim and Gladyshev, 2005). The resolving Cys is generally located in a CGWP motif present in the N-terminal part of the protein (Olry et al., 2002; Tarrago et al., 2009b). In the case of mammalian MsrBs, one or two resolving Cys, located in the N-terminal extremity of the protein, can be involved in the regeneration process (Cao et al., 2018; Kim and Gladyshev, 2005). Of note, most MsrBs possess two CX_2C clusters coordinating of a structural Zn atom (Shumilina et al., 2014). Here, we analyzed the 651 complete fungal MsrB protein sequences. Like MsrAs, they consisted of a single domain and ranged in length from 95 to 289 amino acids (Supplementary data 2A).

Most of the variations in size were due to the presence of an N-terminal extension potentially involved in subcellular targeting. The majority of fungal MsrBs were predicted to be addressed to the mitochondria (61 %) (Figure 2B & Supplementary data 2A). The other proteins were either predicted to be localized in the cytosol (35 %), secreted (1 %), targeted to another compartment, or were not clearly predicted to be addressed to a subcellular compartment (6 %) (Figure 2B & Supplementary data 2A). Almost all fungal MsrBs (> 99 %) possessed the features of the *S. cerevisiae* enzyme (Figure 3B & Supplementary data 2B): i) the two CX_2C motifs involved in the coordination of a Zn atom, ii) the resolving Cys^{97} (from *S. cerevisiae* MsrB residue numbering) included in a $^{97}CGW^{99}$ motif, iii) the conserved His^{134} and His^{136} implicated in substrate binding, together with Arg^{155} and Asn^{159} (Lowther et al., 2002; Neiers et al., 2007), and iv) the catalytic Cys^{157} located in the $^{155}RXCXN^{159}$ motif. The Gly^{98}, Gly^{135}, Val^{137}, Phe^{138} and Glu^{141} residues were also conserved in all these canonical fungal MsrBs (Figure 3B & Supplementary data 2B). Only four sequences (< 1 %) remarkably differed from these common features (Supplementary data 2C), which were identified inthe Orbiliomycetes *Arthrobotrys oligospora* and *Monacrosporium haptotylum*, the Taphrinomycotina *Protomyces lactucaedebilis* and the Chytridiomycetes *Blyttiomyces helices*. These MsrBs lacked the resolving Cys at position 97, which was substituted by a Ser or a Thr. A similar feature was found in MsrBs that either use an alternative Cys for the resolution of the sulfenic acid, such as the human MsrB3 (Cao et al., 2018), or do not possess a resolving Cys, such as the *A. thaliana* MsrB1 for which the sulfenic acid is directly reduced by external reductants (Tarrago et al., 2010, 2009b). Of note, the *Blyttiomyces helices* MsrB lacked any other Cys to act as alternative resolving residue. Finally, another unusual feature was the substitution of the Arg^{155}, strictly conserved in all the MsrBs identified in fungi, and in all yet described MsrBs, by an His in the MsrB from *Protomyces lactucaedebilis* (Supplementary data 2C). Interestingly, searching for MsrB having this atypical HYCIN motif, we found 50 sequences, mainly from poorly characterized bacteria and archaea, but also from the pathogenic worm Loa Loa (Supplementary data 2D). The conservation of this unusual motif at the active site in various organisms and the similarity between His and Arg residues suggest that these MsrBs could have conserved their catalytic activity.

Very few fRMsrs have been characterized so far. However, sequence comparison studies indicated that the catalytic Cys is located in the HIAC motif situated in the middle of the protein sequence and that the two resolving Cys are located ∼30 and ∼40 amino acids upstream in the N-terminal direction (Gruez et al., 2010; Kwak et al., 2010; Le et al., 2009; Lin et al., 2007). The 589 full length fRMsr sequences analyzed here had a single fRMsr domain. Their length varied from 77 to 394 amino acids, with variations in the size of the N-terminal parts (Supplementary data 3A). Most of the proteins were predicted to be localized in the cytoplasm (Figure 2C & Supplementary data 3A). A few proteins were predicted to be secreted (1 %), targeted the mitochondria (1%), to other compartments or without prediction for subcellular targeting (3 %) (Figure 2C & Supplementary data 3A). Similarly to MsrAs and MsrBs, fRMsr showed a strong

conservation of the sequence features of the *S. cerevisiae* protein. Almost all (> 99 %) sequences possessed the resolving Cys in positions 91 and 101 (according to *S. cerevisiae* fRMsr residue numbering) and the catalytic Cys^{125} included in a $^{122}H[I/V]XCD^{126}$ motif (Figure 3C & Supplementary data 3B). We also observed the strict conservation of the Trp^{66}, Tyr^{70}, Glu^{132}, Asp^{149} and Asp^{151}, previously shown to be involved in substrate binding and catalysis. Other important residues involved in substrate binding and catalysis (Gruez et al., 2010) were also conserved or substituted by residues with similar properties in positions 82, 85, 94, 123, 129 and 132. The Pro^{84}, Gly^{87}, His^{122} and Val^{134} were also strictly conserved in these canonical fungal fRMsrs (Figure 3C & Supplementary data 3B). Only three fRMsr sequences from the Agaricomycetes *Scleroderma citrinum*, *Dendrothele bispora* and *Pisolithus tinctorius* presented peculiar characteristics. The first two lacked the potential resolving Cys^{101} and may be still able to reduce the free MetO, but in the latest, the catalytic Cys^{125} was substituted by an Arg, likely precluding catalytic activity (Supplementary data 3C).

Altogether, these results highlight potential subcellular targeting of the fungal Msrs, uncover few proteins with unusual sequence characteristics, but principally showed that the protein sequences of each type of Msr are globally very well conserved throughout the fungal kingdom regarding their Cys contents and the conservation of other residues involved in catalysis.

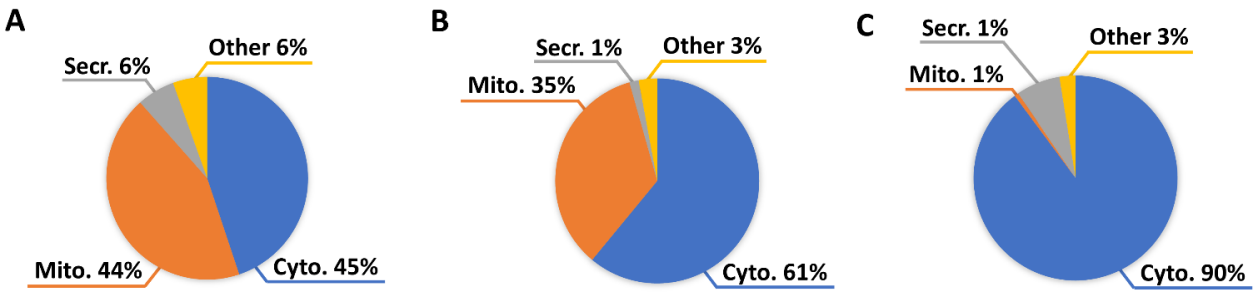


Figure 2 Potential subcellular targeting of fungal Msrs. In the circle chart of subcellular prediction for fungal MsrA (**A**), MsrB (**B**) and fRMsr (**C**), the proportion of proteins predicted to be localized in the cytosol (‘*Cyto.’*), to be addressed the mitochondria (‘*Mito.’*) or secreted (‘*Secr.’*) are shown. The label *‘Other’* indicates the proportion of proteins predicted to be addressed to other compartments or for which no prediction was made.

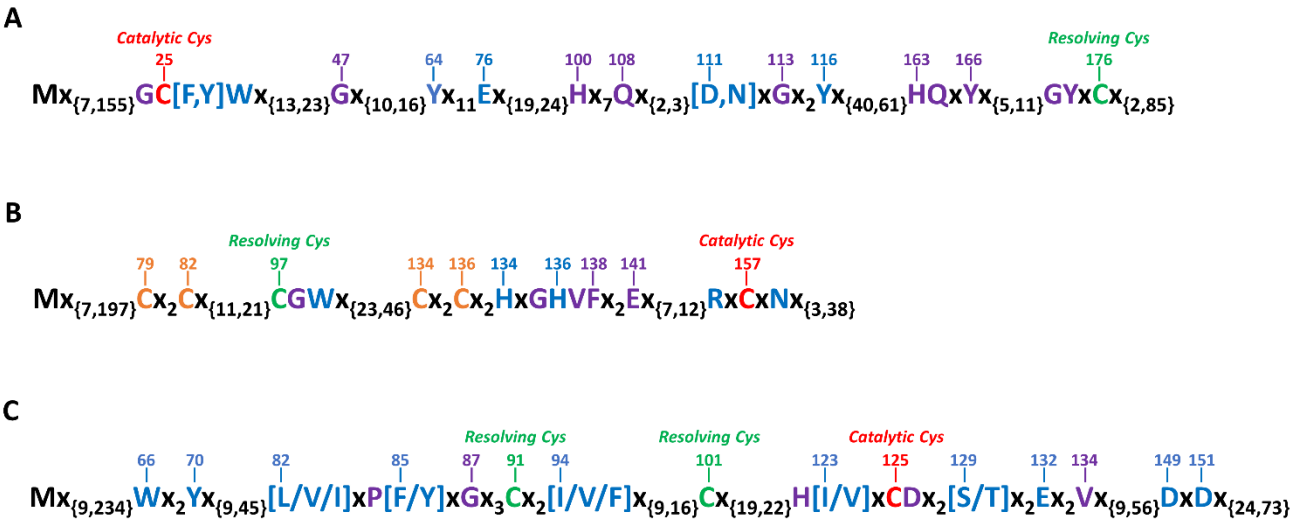


Figure 3 Protein sequence characteristics of canonical fungal Msrs. In this representation of canonical MsrAs (A), MsrBs, (B) and fRMsrs (C), the catalytic Cys (in *red*), the resolving Cys (in *green*) and the residues previously shown to be involved in catalysis and/or substrate binding (in *blue*) are shown. The residues in *purple* are conserved in all canonical fungal Msrs. In B, the Cys residues labeled in *orange* correspond to Zn binding residues. The numberings are based on *S. cerevisiae* MsrA (A), MsrB (B) and fRMsr (C).

Phylogenetic analysis of fungal Msrs and indications for horizontal gene transfers

The phylogenetic relationship of fungal MsrAs globally matched the expected clustering for early-diverging fungi, Ascomycota and Basidiomycota sequences (Figure 4). Surprisingly however, a few Ascomycota MsrA sequences clustered with Basidiomycota sequences (indicated by an asterisk on Figure 4A). We also noticed the clustering all the MsrB sequences from Pucciniomycotina (Basidiomycetes) species with Ascomycota sequences (Figure 4B). Strinkingly, the MsrB sequences from early-diverging fungi did not group in a single cluster and were interspersed in clusters containing Basidiomycota or Ascomycota sequences. For fRMsrs, we observed three distinct clusters containing the protein sequences from Basidiomycota, Ascomycota and early-diverging fungi, respectively (Figure 4C). However, two sequences from early-diverging fungi were included in the cluster containing the Ascomycota sequences (Figure 4C). Altogether, these results showed that, overall, the phylogeny of Msrs was congruent with the phylogeny of the fungi (Delaye et al., 2007; Zhang and Weissbach, 2008).

However, several discrepancies suggested that some Msrs could have arised from horizontal gene transfers. These discrepancies were: i) the positioning in phylogenetic clusters not reflecting the phylogeny of the species from which they were isolated (Figure 4), ii) the presence non-canonical sequence features (Figure 5), and iii) the presence in strictly anaerobic organisms in the case of the *fRmsr* genes present in three Neocallimastigomycetes genomes, whereas other organisms from the same phylum had no *msr* genes (Table 1). To evaluate the possibility

of horizontal gene transfers, we selected all Msr sequences (67 MsrAs, 48 MsrBs and 8 fRMsr) with one or more of these discrepancies and searched for their closest putative homologs by BLASTP search in all organisms recorded in the NCBI nr database. We discarded all the sequences for which the closest homologs were found in fungi. Indeed, because of the strong conservation the protein sequences, it would be difficult to ascertain the occurrence of fungus-to-fungus horizontal gene transfers. We retrieved 17 MsrA, three MsrB and three fRMsr protein sequences for further analysis (Figure 5). Except for two MsrBs from Orbiliomycetes (Ascomycota), all these Msr sequences were from early-diverging fungi (Table 2). For each type of Msr, we used the selected protein sequences, together with the respective 25 sequences with the highest identity score from fungi one the one hand and from non-fungal organisms on the other hand for phylogenetic analysis (Figure 5). For MsrAs, a potential horizontal gene transfer was observed for two selenocysteine-containing MsrAs from *Gonapodya prolifera* that grouped together with bacterial and amoeba MsrAs, in a cluster devoid of any other fungal MsrA (Figure 5A). The difference in GC content between a gene acquired by horizontal gene transfer and its genomic environment has been proposed as a strong indicator of such an event (Ravenhall et al., 2015). We compared the GC content of the *Gonapodya prolifera* MsrA coding sequences to the GC content of the complete scaffold containing the gene. For both genes, the GC content of the MsrA coding sequence was higher than that of the corresponding scaffold (Table 2). This could allow to propose that the genes coding for these two MsrA arose from horizontal gene transfer, potentially from the bacteria *Alkaliphilus oremlandii,* which has the closest selenocysteine-containing MsrA homolog we found (Kim et al., 2009), or from a close relative species (Figure 5C). In addition, the homology between the two *Gonapodya prolifera* MsrA genes suggested they arose from one horizontal gene transfer event followed by a gene duplication (Figure 5A).

In the case of MsrBs, the three selected sequences clustered with bacterial MsrBs (Supplementary Figure 2). For all three, the GC contents of the coding sequence was higher than that of the scaffold in which the gene was present (Table 2). Two MsrBs from the Orbiliomycetes species *Arthrobotrys oligospora and Monacrosporium haptotylum* had MsrBs from the bacteria *Sphingomonas* as closest homologs, whereas the MsrB from the Taphrinomycotina *Protomyces lactucaedebilis* had MsrBs from the cyanobacteria *Calothrix parasitica* as closest homologs. Intriguinly, the MsrB from *P. lactucaedebilis* has anunusual HYCIN active site, whereas its closest homolog in *C. parasitica* has a canonical active site. This indicated that these three MsrBs could have arisen by horizontal gene transfer from a *Sphingomonas* bacteria and a cyanobacteria, respectively.

The three fRMsrs we selected as candidates to horizontal gene transfer were from the Neocallimastigomycetes species *Anaeromyces robustus, Neocallimastix californiae and Orpinomyces sp.* (Table 1). In our phylogenetic analysis, these three fRMsr sequences clustered with bacterial sequences (Figure 5C). Moreover, the GC content of their coding sequence was

higher than that of the scaffolds in which the genes were located (Table 2). Of note, two other fungal fRMsr sequences, one from *Piromyces* sp. and one from *Mortierella* sp. GBA39, were also present in the same cluster of sequences. These two sequences were omitted from our genomic search because of the absence of the corresponding genomes in the MycoCosm database. Altogether, these results strongly argue for the fact that the presence of the *fRmsr* in the genomes of these anaerobic fungi arose from horizontal gene transfers from firmicute bacteria (Table 2).

Altogether, these phylogenetic analyses showed that, although the phylogeny of Msrs respected the phylogeny of the fungi, few horizontal gene transfer events occurred in each Msr family.

Table 2 Potential horizontal *msr* gene transfers from bacteria to fungi

a For both Gonapodya prolifera selenocysteine-containing MsrAs, the MycoCosm accession numbers presented here refer to only a part of the proteins.

b The horizontal gene transfers of Anaeromyces robustus and Piromyces sp. fRmsr genes have been shown recently (Murphy et al., 2019).

c The presented protein accessions are from NCBI (https://www.ncbi.nlm.nih.gov/protein).

n.d, not determined.

Fungal group	Genome	Protein accession	Number of exons	CDS GC content (%)	Scaffold GC content (%)	Organisms with closest Msr homolog
MsrA						
25	*Gonapodya prolifera v1.0*	135492[a]	3	56.4	52.5	*Alkaliphilus oremlandii* OhILAs
25	*Gonapodya prolifera v1.0*	159800/ 159692[a]	4	58.9	53.9	*Alkaliphilus oremlandii* OhILAs
MsrB						
8	*Arthrobotrys oligospora ATCC 24927*	9001	1	55.6	44.4	*Sphingomonas* bacterium
8	*Monacrosporium haptotylum CBS 200.50*	10089	1	55.8	46.7	*Sphingomonas* bacterium
16	*Protomyces lactucaedebilis 12-1054 v1.0*	391028	1	54.9	51.2	*Calothrix parasitica*
fRmsr						
26	*Anaeromyces robustus v1.0*	328892[b]	1	30.1	15.6	*Pseudobutyrivibrio* sp. LB2011
26	*Neocallimastix californiae G1 v1.0*	697339	1	27.5	16.5	*Erysipelotrichaceae* bacterium
26	*Orpinomyces sp.*	1191427	1	34.8	16.8	*Pseudobutyrivibrio ruminis*
26	*Piromyces sp.*	AWI66787.1[b,c]	1	31.3	n.d.	*Pseudobutyrivibrio ruminis*
19	*Mortierella sp. GBA39*	KAF9144191.1[c]	1	51	55.8	*Paenibacillus rhizosphaerae*

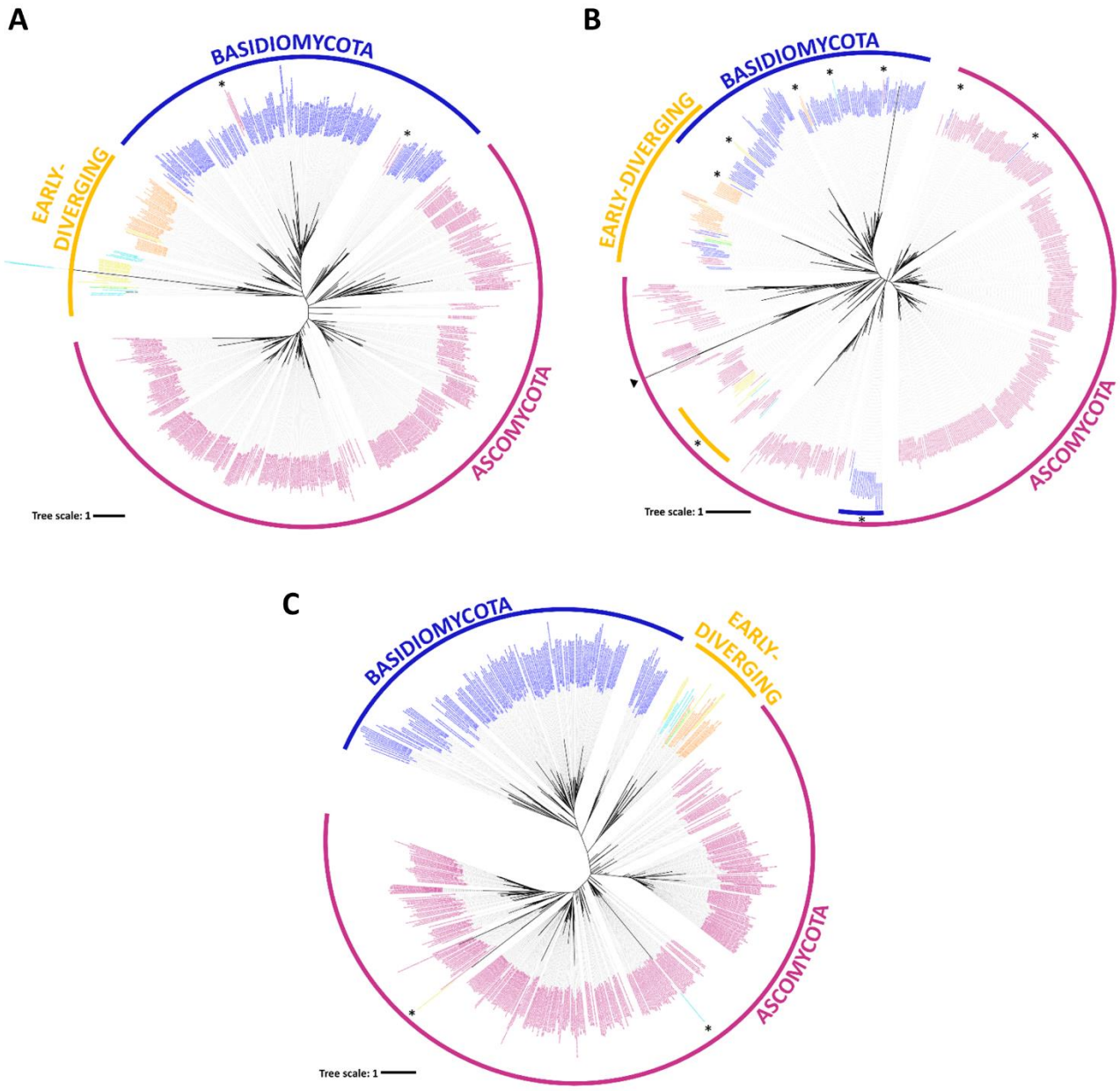


Figure 4 RAxML Phylogenetic trees of fungal MsrA (**A)** MsrB (**B)** and fRMsrs (**C)**. Sequences clustering outside of their taxonomical group are indicated with stars.

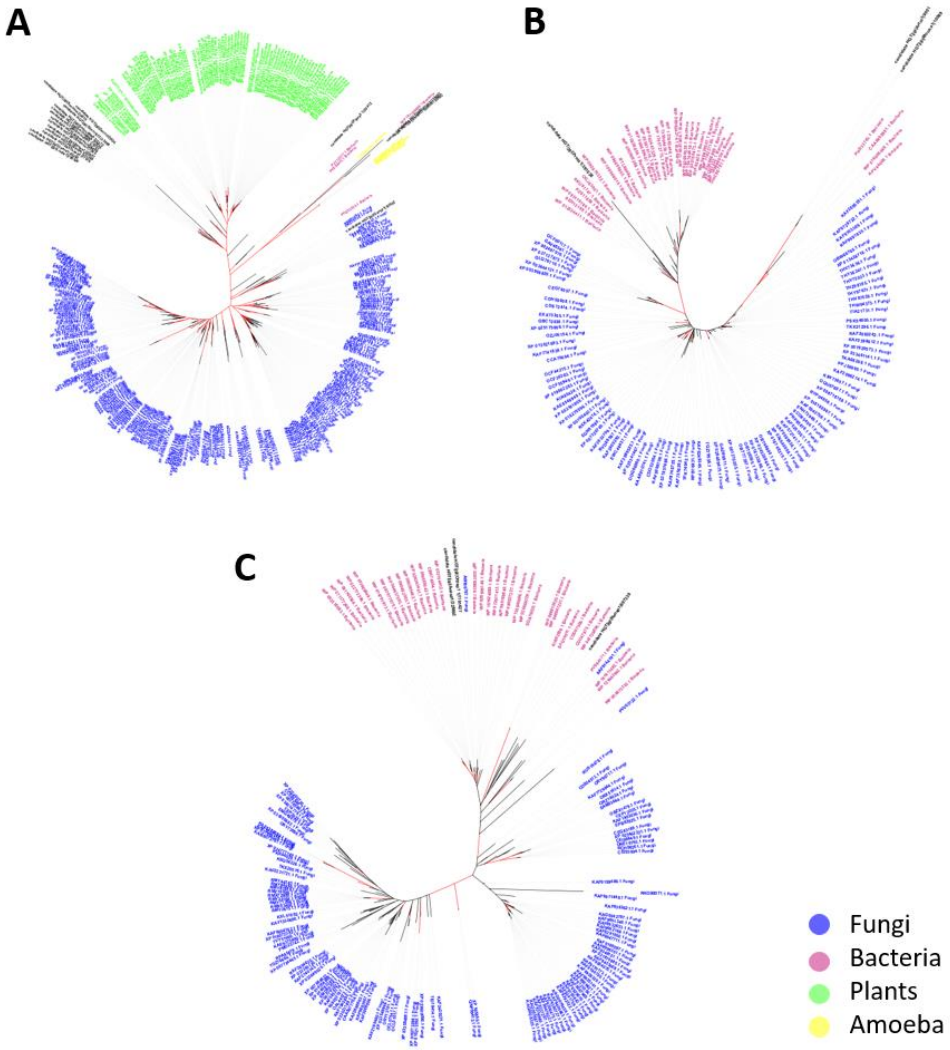


Figure 5 Phylogenetic analysis of fungal Msr tested for horizontal gene transfer (in black). The Msr sequences from fungi, plants, amoeba and bacteria are in *blue*, *green*, *yellow* and *purple*, respectively. Branches with bootstrap values over 70 are in *red*. The phylogenetic tree was built with RAxML v. 8.2 and represented using iTOL (https://itol.embl.de/) (Letunic and Bork, 2019). **A.** MsrAs candidates to HGT are from: *Piptocephalis cylindrospora* RSA 2659 single-cell v3.0 (20412); *Syncephalis pseudoplumigaleata* Benny S71-1 single-cell v1.0 (33365; 23947); *Thamnocephalis sphaerospora* RSA 1356 single-cell v1.0 (27979); *Coemansia reversa* NRRL 1564 v1.0 (43282); *Dimargaris cristalligena* RSA 468 single-cell v1.0 (33956); *Linderina pennispora* ATCC 12442 v1.0 (280520); *Smittium culicis* GSMNP (5372; 7035); *Smittium culicis* ID-206-W2 (5101; 8129); *Smittium mucronatum* ALG-7-W6 (4510; 6899; 7265); *Zancudomyces culisetae* COL-18-3 (924) and *Gonapodya prolifera* v1.0 (135492; 159800/159692; 57902). **B.** MsrBs candidates to HGT are from: *Arthrobotrys oligospora* ATCC 24927 (9001); *Monacrosporium haptotylum* CBS 200.50 (10089) and *Protomyces lactucaedebilis* 12-1054 v1.0 (391028). **C.** FRMsrs candidates to HGT are from: *Anaeromyces robustus* v1.0 (328892); *Neocallimastix californiae* G1 v1.0 (697339) and *Orpinomyces sp* (1191427).

Discussion

Despite the importance of fungi, the roles of Met oxidation and the *msr* gene families in these eukaryotes have been barely addressed. In this article, we took advantage of the exploding numbers of publicly available fungal genomes to perform a global search for *msr* genes in nearly 700 fungal genomes covering all the fungal diversity (Grigoriev et al., 2014) (Figure 1). Overall, we found that most fungal genomes contain one gene coding for each thiol-oxidoreductase type, i.e. MsrA, MsrB and fRMsr. Despite the presence of molybdenum-containing enzymes in some fungi (Peng et al., 2018), we did not find any homologs of the bacterial molybdenum-containing Msr in our search.

The phylogenetic analyses and inspection of protein sequence features revealed that each type of thiol-oxidoreductase Msr is globally very well conserved across the fungal kingdom (Figures 3 and 4). This is consistent with the prokaryotic origin of these genes (Delaye et al., 2007; Le et al., 2009; Lin et al., 2007; Zhang and Weissbach, 2008). However, the identification of *fRmsr* genes in almost all genomes across the fungal kingdom was surprising (Figure 1). Indeed, no *fRmsr* genes were identified so far from multicellular eukaryotes (Le et al., 2009). Very likely, the most obvious reason was the number of eukaryotic genomes considered in this study (160), and the smaller number of fungal genomes available at that time (Grigoriev et al., 2014; Le et al., 2009). Because the number of genomes now available for multicellular eukaryote organisms has increased dramatically in the last years, we took the opportunity of this study to reevaluate the presence of *fRmsr* genes in other multicellular eukaryotes. We searched for *fRmsr* genes in the plant (Viridiplantae) and animal genomes available in the nr database of NCBI by BLAST search. After removing the hits corresponding to unicellular organisms (mostly green algae) and bacterial contaminants, we found only a handful of multicellular eukaryotes apparently possessing a gene coding for a fRMsr (Supplementary Table 1). These organisms, a plant and a few insects, are not phylogenetically related together, indicating that the presence of the gene in their genomes may result from horizontal gene transfers, and is not related to the conservation of the gene in their lineage. This confirms that the *fRmsr* gene is present in multicellular eukaryotes but is mostly restricted to the fungal kingdom.

We observed that Msr sequences are strongly conserved across the fungal kingdom. Unusual features were found in the active sites of a few MsrA and MsrB, that are also found in some non-fungal organisms. It would be interesting to test the biochemical properties of non-canonical enzymes (e.g. substrate specificity).

Interestingly, we identified two MsrA genes in *Gonapodya prolifera* that could have been acquired via HGT. The two genes have their closest homologs in *Alkaliphilus oremlandii*, a strict anaerobic bacterium phylogenetically associated with the Firmicutes. Bacteria from the genus *Alkaliphilus* are found in sediments and ponds, and some strains are able to grow in toxic environments. For example, strains from *A. oremlandii* and *Alkaliphilus metalliredigens* can

respire arsenate and thiosulfate or grow in toxic metal-contaminated sites, respectively (Fischer et al., 2008; Hwang et al., 2016). Similarly, *Gonapodya prolifera* occurs on fruits submerged into ponds with low levels of oxygen (Fuller and Clay, 1993). The presence of the fungus and the bacteria in a same ecological niche could have favored the horizontal gene transfer.

We also noticed that MsrB gene transfers occurred in two Ascomycota species from the Orbiliomycetes family, *Arthrobotrys oligospora and Monacrosporium haptotylum.* Both fungi are renowned for their ability to trap nematodes in soil (Niu and Zhang, 2011; Meerupati et al., 2013). We identified sphingomonads as potential donors for the MsrB gene. Sphingomonads have been isolated from many different land and water habitats, and in environments contaminated with toxic compounds (Nishiyama et al., 1992). The co-ocurence in the same habitat of a sphingomonad donor and a common ancestor of *A. oligospora* and *M. haptotylum* is then plausible. Regarding MsrB, another potential HGT was identified, from a donor cyanobacteria close to the marine *Calothrix parasitica* to the plant pathogen fission yeast *Protomyces lactucaedebilis* (Taphrinomycotina). In this case however, our current knowledge on the ecology of these organisms does not support the co-occurrence of the donor and acceptor organisms in a same ecological niche.

Finally, we identified *fRmsr* HGT events from Firmicutes bacteria to the Neocallimastigomycetes species *Anaeromyces robustus, Neocallimastix californiae, Orpinomyces* sp. and *Piromyces* sp. Interestingly, the four species of fungi considered here are living in the anaerobic gut of ruminants, and the bacteria with the closest homologs are firmicutes, and also members of the ruminant gut microbiota (Murphy et al., 2019). Moreover, the horizontal gene transfer of *fRmsr* genes from firmicutes to *Anaeromyces robustus* and *Piromyces* sp. have been shown recently (Murphy et al., 2019). Altogether, these results strongly argue for the fact that the presence of the *fRmsr* in the genomes of these anaerobic fungi arose from horizontal gene transfers from firmicute.

Material and methods

MsrA, MsrB and fRMsr protein sequences from *Saccharomyces cerevisiae* were used as BLASTP and TBLASTN queries to identify msr genes in 682 published genomes available in the MycoCosm database.

For each type of msr, we built a multiple sequence alignment with all the identified sequences using MAFFT version v7.429. Each alignment was then trimmed to remove poorly aligned regions using trimAl 1.2 and manually inspected for conservation of the catalytic residues. A phylogenetic tree was finally constructed for each type of msr using RAXML Master Pthread version 8.2.12 (PROTGAMMAWAG model and 500 bootstrap). The phylogenetic tree was represented using iTOL (https://itol.embl.de/) (Letunic and Bork, 2019) and edited in Inkscape.

To check for potential horizontal gene transfers, the “non-canonical” msr sequences along with duplicated msr genes distantly located on the phylogenetic trees were retrieved. Those sequences were further used as blast queries against the non-redundant protein sequence (nr) database and the best 100 hits were retrieved. The blast results were manually inspected and the fungal sequences for which the highest identity scores were obtained in organisms outside the fungal kingdom were retrieved. Those genes were further used as blast queries against nr excluding the taxa Fungi (taxid:4751) and the 25 hits with the highest identity scores were retrieved. In parallel, the same genes were used as queries against the Fnugal nr database and the best 50 hits with the highest identity scores were retrieved, excluding those with 100% identity with the query. The retrieved fungal hits were manually inspected to exclude “non-canonical sequences”. Finally, for each msr type, a phylogenetic tree was constructed as explained above, including the candidate gene for HGT, plus 25 non-fungal and 25 fungal with high identity scores.

References

Andrieu, C., Vergnes, A., Loiseau, L., Aussel, L., Ezraty, B., 2020. Characterisation of the periplasmic methionine sulfoxide reductase (MsrP) from Salmonella Typhimurium. Free Radic. Biol. Med. 160, 506–512. https://doi.org/10.1016/j.freeradbiomed.2020.06.031

Antoine, M., Gand, A., Boschi-Muller, S., Branlant, G., 2006. Characterization of the Amino Acids from Neisseria meningitidis MsrA Involved in the Chemical Catalysis of the Methionine Sulfoxide Reduction Step. J. Biol. Chem. 281, 39062–39070. https://doi.org/10.1074/jbc.M608844200

Boschi-Muller, S., Azza, S., Sanglier-Cianferani, S., Talfournier, F., Van Dorsselear, A., Branlant, G., 2000. A sulfenic acid enzyme intermediate is involved in the catalytic mechanism of peptide methionine sulfoxide reductase from Escherichia coli. J. Biol. Chem. 275, 35908–35913. https://doi.org/10.1074/jbc.M006137200

Breitenbach, M., Weber, M., Rinnerthaler, M., Karl, T., Breitenbach-Koller, L., 2015. Oxidative Stress in Fungi: Its Function in Signal Transduction, Interaction with Plant Hosts, and Lignocellulose Degradation. Biomolecules 5, 318–342. https://doi.org/10.3390/biom5020318

Brot, N., Weissbach, L., Werth, J., Weissbach, H., 1981. Enzymatic reduction of protein-bound methionine sulfoxide. Proc. Natl. Acad. Sci. U. S. A. 78, 2155–2158.

Cao, Z., Mitchell, L., Hsia, O., Scarpa, M., Caldwell, S.T., Alfred, A.D., Gennaris, A., Collet, J.-F., Hartley, R.C., Bulleid, N.J., 2018. Methionine sulfoxide reductase B3 requires resolving cysteine residues for full activity and can act as a stereospecific methionine oxidase. Biochem. J. 475, 827–838. https://doi.org/10.1042/BCJ20170929

Delaye, L., Becerra, A., Orgel, L., Lazcano, A., 2007. Molecular evolution of peptide methionine sulfoxide reductases (MsrA and MsrB): on the early development of a mechanism that protects against oxidative damage. J. Mol. Evol. 64, 15–32. https://doi.org/10.1007/s00239-005-0281-2

Dhouib, R., Othman, D.S.M.P., Lin, V., Lai, X.J., Wijesinghe, H.G.S., Essilfie, A.-T., Davis, A., Nasreen, M., Bernhardt, P.V., Hansbro, P.M., McEwan, A.G., Kappler, U., 2016. A Novel, Molybdenum-Containing Methionine Sulfoxide Reductase Supports Survival of Haemophilus influenzae in an In vivo Model of Infection. Front. Microbiol. 7, 1743. https://doi.org/10.3389/fmicb.2016.01743

Drazic, A., Miura, H., Peschek, J., Le, Y., Bach, N.C., Kriehuber, T., Winter, J., 2013. Methionine oxidation activates a transcription factor in response to oxidative stress. Proc. Natl. Acad. Sci. U. S. A. 110, 9493–9498. https://doi.org/10.1073/pnas.1300578110

Erickson, J.R., Joiner, M.A., Guan, X., Kutschke, W., Yang, J., Oddis, C.V., Bartlett, R.K., Lowe, J.S., O'Donnell, S.E., Aykin-Burns, N., Zimmerman, M.C., Zimmerman, K., Ham, A.-J.L., Weiss, R.M., Spitz, D.R., Shea, M.A., Colbran, R.J., Mohler, P.J., Anderson, M.E., 2008. A dynamic pathway for calcium-independent activation of CaMKII by methionine oxidation. Cell 133, 462–474. https://doi.org/10.1016/j.cell.2008.02.048

Ezraty, B., Bos, J., Barras, F., Aussel, L., 2005. Methionine sulfoxide reduction and assimilation in Escherichia coli: new role for the biotin sulfoxide reductase BisC. J. Bacteriol. 187, 231–237. https://doi.org/10.1128/JB.187.1.231-237.2005

Fisher, E., Dawson, A.M., Polshyna, G., Lisak, J., Crable, B., Perera, E., Ranganathan, M., Thangavelu, M., Basu, P. and Stolz, J.F. 2008. Transformation of Inorganic and Organic Arsenic byAlkaliphilus oremlandiisp. nov. Strain OhILAs. Annals of the New York Academy of Sciences, 1125: 230-241. https://doi.org/10.1196/annals.1419.006

Fuller, Melvin S., and Ronald P. Clay. 2021. Observations of Gonapodya in Pure Culture: Growth, Development and Cell Wall Characterization. Mycologia, vol. 85, no. 1, 1993, pp. 38–45. JSTOR, www.jstor.org/stable/3760475. Accessed 18 Jan.

Gallmetzer, A., Silvestrini, L., Schinko, T., Gesslbauer, B., Hortschansky, P., Dattenböck, C., Muro-Pastor, M.I., Kungl, A., Brakhage, A.A., Scazzocchio, C., Strauss, J., 2015. Reversible Oxidation of a Conserved Methionine in the Nuclear Export Sequence Determines Subcellular Distribution and Activity of the Fungal Nitrate Regulator NirA. PLoS Genet. 11, e1005297. https://doi.org/10.1371/journal.pgen.1005297

Gennaris, A., Ezraty, B., Henry, C., Agrebi, R., Vergnes, A., Oheix, E., Bos, J., Leverrier, P., Espinosa, L., Szewczyk, J., Vertommen, D., Iranzo, O., Collet, J.-F., Barras, F., 2015. Repairing oxidized proteins in the bacterial envelope using respiratory chain electrons. Nature 528, 409–412. https://doi.org/10.1038/nature15764

Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., Shabalov, I., 2014. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. 42, D699–D704. https://doi.org/10.1093/nar/gkt1183

Grimaud, R., Ezraty, B., Mitchell, J.K., Lafitte, D., Briand, C., Derrick, P.J., Barras, F., 2001. Repair of oxidized proteins. Identification of a new methionine sulfoxide reductase. J. Biol. Chem. 276, 48915–48920. https://doi.org/10.1074/jbc.M105509200

Gruez, A., Libiad, M., Boschi-Muller, S., Branlant, G., 2010. Structural and biochemical characterization of free methionine-R-sulfoxide reductase from Neisseria meningitidis. J. Biol. Chem. 285, 25033–25043. https://doi.org/10.1074/jbc.M110.134528

Hwang, C., Copeland, A., Lucas, S., Lapidus, A., Barry, K., Detter, J. C., Glavina Del Rio, T., Hammon, N., Israni, S., Dalin, E., Tice, H., Pitluck, S., Chertkov, O., Brettin, T., Bruce, D., Han, C., Schmutz, J., Larimer, F., Land, M. L., Hauser, L., … Fields, M. W. (2016). Complete Genome Sequence of Alkaliphilus metalliredigens Strain QYMF, an Alkaliphilic and Metal-Reducing Bacterium Isolated from Borax-Contaminated Leachate Ponds. Genome announcements, 4(6), e01226-16. https://doi.org/10.1128/genomeA.01226-16

James, T.Y., Stajich, J.E., Hittinger, C.T., Rokas, A., 2020. Toward a Fully Resolved Fungal Tree of Life. Annu. Rev. Microbiol. 74, 291–313. https://doi.org/10.1146/annurev-micro-022020-051835

Jiang, B., Moskovitz, J., 2018. The Functions of the Mammalian Methionine Sulfoxide Reductase System and Related Diseases. Antioxid. Basel Switz. 7. https://doi.org/10.3390/antiox7090122

Jiang, G., Zeng, J., Li, Z., Song, Y., Yan, H., He, J., Jiang, Y., Duan, X., 2020. Redox regulation of the NOR transcription factor is involved in the regulation of fruit ripening in tomato. Plant Physiol. https://doi.org/10.1104/pp.20.00070

Jo, H., Cho, Y.-W., Ji, S.-Y., Kang, G.-Y., Lim, C.-J., 2014. Protective roles of methionine-R-sulfoxide reductase against stresses in Schizosaccharomyces pombe. J. Basic Microbiol. 54, 72–80. https://doi.org/10.1002/jobm.201200397

Kachurin, A.M., Golubev, A.M., Geisow, M.M., Veselkina, O.S., Isaeva-Ivanova, L.S., Neustroev, K.N., 1995. Role of methionine in the active site of alpha-galactosidase from Trichoderma reesei. Biochem. J. 308, 955–964.

Katinka, M.D., Duprat, S., Cornillot, E., Méténier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretaillade, E., Brottier, P., Wincker, P., Delbac, F., El Alaoui, H., Peyret, P., Saurin, W., Gouy, M., Weissenbach, J., Vivarès, C.P., 2001. Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi. Nature 414, 450–453. https://doi.org/10.1038/35106579

Kato, M., Yang, Y.-S., Sutter, B.M., Wang, Y., McKnight, S.L., Tu, B.P., 2019. Redox State Controls Phase Separation of the Yeast Ataxin-2 Protein via Reversible Oxidation of Its Methionine-Rich Low-Complexity Domain. Cell 177, 711-721.e8. https://doi.org/10.1016/j.cell.2019.02.044

Kaya, A., Koc, A., Lee, B.C., Fomenko, D.E., Rederstorff, M., Krol, A., Lescure, A., Gladyshev, V.N., 2010. Compartmentalization and regulation of mitochondrial function by methionine sulfoxide reductases in yeast. Biochemistry 49, 8618–8625. https://doi.org/10.1021/bi100908v

Kim, H.-Y., 2012. Glutaredoxin serves as a reductant for methionine sulfoxide reductases with or without resolving cysteine. Acta Biochim. Biophys. Sin. 44, 623–627. https://doi.org/10.1093/abbs/gms038

Kim, H.-Y., Fomenko, D.E., Yoon, Y.-E., Gladyshev, V.N., 2006. Catalytic advantages provided by selenocysteine in methionine-S-sulfoxide reductases. Biochemistry 45, 13697–13704. https://doi.org/10.1021/bi0611614

Kim, H.-Y., Gladyshev, V.N., 2005. Different catalytic mechanisms in mammalian selenocysteine- and cysteine-containing methionine-R-sulfoxide reductases. PLoS Biol. 3, e375. https://doi.org/10.1371/journal.pbio.0030375

Kim, H.-Y., Zhang, Y., Lee, B.C., Kim, J.-R., Gladyshev, V.N., 2009. The selenoproteome of Clostridium sp. OhILAs: characterization of anaerobic bacterial selenoprotein methionine sulfoxide reductase A. Proteins 74, 1008–1017. https://doi.org/10.1002/prot.22212

Koc, A., Gasch, A.P., Rutherford, J.C., Kim, H.-Y., Gladyshev, V.N., 2004. Methionine sulfoxide reductase regulation of yeast lifespan reveals reactive oxygen species-dependent and -independent

components of aging. Proc. Natl. Acad. Sci. U. S. A. 101, 7999–8004. https://doi.org/10.1073/pnas.0307929101

Kryukov, G.V., Kumar, R.A., Koc, A., Sun, Z., Gladyshev, V.N., 2002. Selenoprotein R is a zinc-containing stereo-specific methionine sulfoxide reductase. Proc. Natl. Acad. Sci. U. S. A. 99, 4245–4250. https://doi.org/10.1073/pnas.072603099

Kwak, G.-H., Kim, M.-J., Kim, H.-Y., 2010. Cysteine-125 is the catalytic residue of Saccharomyces cerevisiae free methionine-R-sulfoxide reductase. Biochem. Biophys. Res. Commun. 395, 412–415. https://doi.org/10.1016/j.bbrc.2010.04.036

Le, D.T., Lee, B.C., Marino, S.M., Zhang, Y., Fomenko, D.E., Kaya, A., Hacioglu, E., Kwak, G.-H., Koc, A., Kim, H.-Y., Gladyshev, V.N., 2009. Functional analysis of free methionine-R-sulfoxide reductase from Saccharomyces cerevisiae. J. Biol. Chem. 284, 4354–4364. https://doi.org/10.1074/jbc.M805891200

Lee, B.C., Lee, H.M., Kim, S., Avanesov, A.S., Lee, A., Chun, B.-H., Vorbruggen, G., Gladyshev, V.N., 2018. Expression of the methionine sulfoxide reductase lost during evolution extends Drosophila lifespan in a methionine-dependent manner. Sci. Rep. 8, 1–11. https://doi.org/10.1038/s41598-017-15090-5

Lee, B.C., Péterfi, Z., Hoffmann, F.W., Moore, R.E., Kaya, A., Avanesov, A., Tarrago, L., Zhou, Y., Weerapana, E., Fomenko, D.E., Hoffmann, P.R., Gladyshev, V.N., 2013. MsrB1 and MICALs regulate actin assembly and macrophage function via reversible stereoselective methionine oxidation. Mol. Cell 51, 397–404. https://doi.org/10.1016/j.molcel.2013.06.019

Lee, E.H., Lee, K., Kwak, G.-H., Park, Y.S., Lee, K.-J., Hwang, K.Y., Kim, H.-Y., 2015. Evidence for the dimerization-mediated catalysis of methionine sulfoxide reductase A from Clostridium oremlandii. PloS One 10, e0131523. https://doi.org/10.1371/journal.pone.0131523

Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. Nucleic acids research, 47(W1), W256–W259. https://doi.org/10.1093/nar/gkz239

Lim, J.M., Kim, G., Levine, R.L., 2018. Methionine in Proteins: It's Not Just for Protein Initiation Anymore. Neurochem. Res. https://doi.org/10.1007/s11064-017-2460-0

Lin, Z., Johnson, L.C., Weissbach, H., Brot, N., Lively, M.O., Lowther, W.T., 2007. Free methionine-(R)-sulfoxide reductase from Escherichia coli reveals a new GAF domain function. Proc. Natl. Acad. Sci. U. S. A. 104, 9597–9602. https://doi.org/10.1073/pnas.0703774104

Lourenço dos Santos, S., Petropoulos, I., Friguet, B., 2018. The Oxidized Protein Repair Enzymes Methionine Sulfoxide Reductases and Their Roles in Protecting against Oxidative Stress, in Ageing and in Regulating Protein Function. Antioxidants 7, 191. https://doi.org/10.3390/antiox7120191

Lowther, W.T., Brot, N., Weissbach, H., Honek, J.F., Matthews, B.W., 2000. Thiol-disulfide exchange is involved in the catalytic mechanism of peptide methionine sulfoxide reductase. Proc. Natl. Acad. Sci. U. S. A. 97, 6463–6468.

Lowther, W.T., Weissbach, H., Etienne, F., Brot, N., Matthews, B.W., 2002. The mirrored methionine sulfoxide reductases of Neisseria gonorrhoeae pilB. Nat. Struct. Biol. 9, 348–352. https://doi.org/10.1038/nsb783

Ma, X.-X., Guo, P.-C., Shi, W.-W., Luo, M., Tan, X.-F., Chen, Y., Zhou, C.-Z., 2011. Structural Plasticity of the Thioredoxin Recognition Site of Yeast Methionine S-Sulfoxide Reductase Mxr1. J. Biol. Chem. 286, 13430–13437. https://doi.org/10.1074/jbc.M110.205161

Mariotti, M., Salinas, G., Gabaldón, T., Gladyshev, V.N., 2019. Utilization of selenocysteine in early-branching fungal phyla. Nat. Microbiol. 4, 759–765. https://doi.org/10.1038/s41564-018-0354-9

Meerupati T, Andersson KM, Friman E, Kumar D, Tunlid A, et al. 2013. Genomic Mechanisms Accounting for the Adaptation to Parasitism in Nematode-Trapping Fungi. PLOS Genetics 9(11): e1003909. https://doi.org/10.1371/journal.pgen.1003909

Meyer, V., Basenko, E.Y., Benz, J.P., Braus, G.H., Caddick, M.X., Csukai, M., de Vries, R.P., Endy, D., Frisvad, J.C., Gunde-Cimerman, N., Haarmann, T., Hadar, Y., Hansen, K., Johnson, R.I., Keller, N.P., Kraševec, N., Mortensen, U.H., Perez, R., Ram, A.F.J., Record, E., Ross, P., Shapaval, V., Steiniger, C., van den Brink, H., van Munster, J., Yarden, O., Wösten, H.A.B., 2020. Growing a circular economy with fungal biotechnology: a white paper. Fungal Biol. Biotechnol. 7, 5. https://doi.org/10.1186/s40694-020-00095-z

Moskovitz, J., Flescher, E., Berlett, B.S., Azare, J., Poston, J.M., Stadtman, E.R., 1998. Overexpression of peptide-methionine sulfoxide reductase in Saccharomyces cerevisiae and human T cells provides them with high resistance to oxidative stress. Proc. Natl. Acad. Sci. U. S. A. 95, 14071–14075.

Murphy, C.L., Youssef, N.H., Hanafy, R.A., Couger, M.B., Stajich, J.E., Wang, Y., Baker, K., Dagar, S.S., Griffith, G.W., Farag, I.F., Callaghan, T.M., Elshahed, M.S., 2019. Horizontal Gene Transfer as an Indispensable Driver for Evolution of Neocallimastigomycota into a Distinct Gut-Dwelling Fungal Lineage. Appl. Environ. Microbiol. 85. https://doi.org/10.1128/AEM.00988-19

Naranjo-Ortiz, M.A., Gabaldón, T., 2019. Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. Biol. Rev. Camb. Philos. Soc. 94, 2101–2137. https://doi.org/10.1111/brv.12550

Neiers, F., Sonkaria, S., Olry, A., Boschi-Muller, S., Branlant, G., 2007. Characterization of the amino acids from Neisseria meningitidis methionine sulfoxide reductase B involved in the chemical catalysis and substrate specificity of the reductase step. J. Biol. Chem. 282, 32397–32405. https://doi.org/10.1074/jbc.M704730200

Nicklow, E.E., Sevier, C.S., 2020. Activity of the yeast cytoplasmic Hsp70 nucleotide-exchange factor Fes1 is regulated by reversible methionine oxidation. J. Biol. Chem. 295, 552–569. https://doi.org/10.1074/jbc.RA119.010125

Nishiyama, M., Senoo, K., Wada, H., & Matsumoto, S. 1992. Identification of soil micro-habitats for growth, death and survival of a bacterium, γ-1, 2, 3, 4, 5, 6-hexachlorocyclohexane-assimilating *Sphingomonas paucimobilis*, by fractionation of soil. FEMS Microbiology Letters, 101(3), 145-150.

Niu, X. M., & Zhang, K. Q. (2011). *Arthrobotrys oligospora*: a model organism for understanding the interaction between fungi and nematodes. Mycology, 2(2), 59-78.

Olry, A., Boschi-Muller, S., Marraud, M., Sanglier-Cianferani, S., Van Dorsselear, A., Branlant, G., 2002. Characterization of the methionine sulfoxide reductase activities of PILB, a probable virulence factor from Neisseria meningitidis. J. Biol. Chem. 277, 12016–12022. https://doi.org/10.1074/jbc.M112350200

Peng, T., Xu, Y., Zhang, Y., 2018. Comparative genomics of molybdenum utilization in prokaryotes and eukaryotes. BMC Genomics 19, 691. https://doi.org/10.1186/s12864-018-5068-0

Ranaivoson, F.M., Neiers, F., Kauffmann, B., Boschi-Muller, S., Branlant, G., Favier, F., 2009. Methionine sulfoxide reductase B displays a high level of flexibility. J. Mol. Biol. 394, 83–93. https://doi.org/10.1016/j.jmb.2009.08.073

Ravenhall, M., Škunca, N., Lassalle, F., Dessimoz, C., 2015. Inferring Horizontal Gene Transfer. PLoS Comput. Biol. 11. https://doi.org/10.1371/journal.pcbi.1004095

Rey, P., Tarrago, L., 2018. Physiological Roles of Plant Methionine Sulfoxide Reductases in Redox Homeostasis and Signaling. Antioxidants 7, 114. https://doi.org/10.3390/antiox7090114

Rouhier, N., Kauffmann, B., Tete-Favier, F., Palladino, P., Gans, P., Branlant, G., Jacquot, J.-P., Boschi-Muller, S., 2007. Functional and structural aspects of poplar cytosolic and plastidial type a methionine sulfoxide reductases. J. Biol. Chem. 282, 3367–3378. https://doi.org/10.1074/jbc.M605007200

Sharov, V.S., Ferrington, D.A., Squier, T.C., Schöneich, C., 1999. Diastereoselective reduction of protein-bound methionine sulfoxide by methionine sulfoxide reductase. FEBS Lett. 455, 247–250.

Sharov, V.S., Schöneich, C., 2000. Diastereoselective protein methionine oxidation by reactive oxygen species and diastereoselective repair by methionine sulfoxide reductase. Free Radic. Biol. Med. 29, 986–994.

Shumilina, E., Dobrovolska, O., Dikiy, A., 2014. Evolution of Structural and Coordination Features Within the Methionine Sulfoxide Reductase B Family, in: Hohmann-Marriott, M.F. (Ed.), The Structural Basis of Biological Energy Generation, Advances in Photosynthesis and Respiration. Springer Netherlands, Dordrecht, pp. 199–215. https://doi.org/10.1007/978-94-017-8742-0_11

Sies, H., Jones, D.P., 2020. Reactive oxygen species (ROS) as pleiotropic physiological signalling agents. Nat. Rev. Mol. Cell Biol. 21, 363–383. https://doi.org/10.1038/s41580-020-0230-3

Singh, V.K., Singh, K., Baum, K., 2018. The Role of Methionine Sulfoxide Reductases in Oxidative Stress Tolerance and Virulence of Staphylococcus aureus and Other Bacteria. Antioxidants 7. https://doi.org/10.3390/antiox7100128

Skalski, J.H., Kottom, T.J., Limper, A.H., 2015. Pathobiology of Pneumocystis pneumonia: life cycle, cell wall and cell signal transduction. FEMS Yeast Res. 15. https://doi.org/10.1093/femsyr/fov046

Soriani, F.M., Kress, M.R., Fagundes de Gouvêa, P., Malavazi, I., Savoldi, M., Gallmetzer, A., Strauss, J., Goldman, M.H.S., Goldman, G.H., 2009. Functional characterization of the Aspergillus nidulans methionine sulfoxide reductases (msrA and msrB). Fungal Genet. Biol. FG B 46, 410–417.

Tarrago, L., Gladyshev, V.N., 2012. Recharging oxidative protein repair: catalysis by methionine sulfoxide reductases towards their amino acid, protein, and model substrates. Biochem. Biokhimiia 77, 1097–1107. https://doi.org/10.1134/S0006297912100021

Tarrago, L., Grosse, S., Lemaire, D., Faure, L., Tribout, M., Siponen, M.I., Kojadinovic-Sirinelli, M., Pignol, D., Arnoux, P., Sabaty, M., 2020. Reduction of Protein Bound Methionine Sulfoxide by a Periplasmic Dimethyl Sulfoxide Reductase. Antioxid. Basel Switz. 9. https://doi.org/10.3390/antiox9070616

Tarrago, L., Grosse, S., Siponen, M.I., Lemaire, D., Alonso, B., Miotello, G., Armengaud, J., Arnoux, P., Pignol, D., Sabaty, M., 2018. Rhodobactersphaeroides methionine sulfoxide reductase P reduces R- and S-diastereomers of methionine sulfoxide from a broad-spectrum of protein substrates. Biochem. J. 475, 3779–3795. https://doi.org/10.1042/BCJ20180706

Tarrago, L., Kaya, A., Weerapana, E., Marino, S.M., Gladyshev, V.N., 2012. Methionine sulfoxide reductases preferentially reduce unfolded oxidized proteins and protect cells from oxidative protein unfolding. J. Biol. Chem. 287, 24448–24459. https://doi.org/10.1074/jbc.M112.374520

Tarrago, L., Laugier, E., Rey, P., 2009a. Protein-repairing methionine sulfoxide reductases in photosynthetic organisms: gene organization, reduction mechanisms, and physiological roles. Mol. Plant 2, 202–217. https://doi.org/10.1093/mp/ssn067

Tarrago, L., Laugier, E., Zaffagnini, M., Marchand, C., Le Maréchal, P., Rouhier, N., Lemaire, S.D., Rey, P., 2009b. Regeneration mechanisms of *Arabidopsis thaliana* methionine sulfoxide reductases B by

glutaredoxins and thioredoxins. J. Biol. Chem. 284, 18963–18971. https://doi.org/10.1074/jbc.M109.015487

Tarrago, L., Laugier, E., Zaffagnini, M., Marchand, C.H., Le Maréchal, P., Lemaire, S.D., Rey, P., 2010. Plant thioredoxin CDSP32 regenerates 1-cys methionine sulfoxide reductase B activity through the direct reduction of sulfenic acid. J. Biol. Chem. 285, 14964–14972. https://doi.org/10.1074/jbc.M110.108373

Yin, C., Zheng, L., Zhu, J., Chen, L., Ma, A., 2015. Enhancing stress tolerance by overexpression of a methionine sulfoxide reductase A (MsrA) gene in Pleurotus ostreatus. Appl. Microbiol. Biotechnol. 99, 3115–3126. https://doi.org/10.1007/s00253-014-6365-4

Zhang, X.-H., Weissbach, H., 2008. Origin and evolution of the protein-repairing enzymes methionine sulphoxide reductases. Biol. Rev. Camb. Philos. Soc. 83, 249–257. https://doi.org/10.1111/j.1469-185X.2008.00042.x

ANNEXE 2: Tandem fungal oxidases gate pathogenesis in major food crops.

Summary of my contribution to this work

Considering the massive population growth that the world is continuously facing, food sufficiency is slowly becoming a major concern that modern societies will need to manage. Thus, one of the solutions is to ensure food security against pathogens that threaten plant health. Among plant pathogens, fungi from the phylum Ascomycota, such as *Magnaporthe* and *Colletotrichum*, are strong devastating agents.

In this study, colleagues from the lab aimed at understanding the role of a copper-radical alcohol oxidase (AlcOx) in plant infection. A role for this enzyme in pathogenicity adds to the previously proposed functions in morphological development or lignin degradation. Using the scripts I developed, we could reveal the co-occurrence of a peroxidase gene in the neighboring of a AlcOx gene, with inverted tandem organization, in *Colletotrichum* and *Magnaporthe* genomes exclusively. In addition, I performed the phylogenetic analysis of those two enzymes and built a species tree of selected pathogenic ascomycete fungi.

A series of functional assays were performed by our group and collaborators. We observed that the Perox enzyme from *Colletotrichum orbiculare* activates the paired AlcOx for the oxidation of long chain alcohols and plant waxes. The co-expression of the AlcOx-Perox enzyme pair in *Colletotrichum* and *Magnaporthe* species during the early penetration phase of plant cell wall, and the drastic reduction in fungal pathogenicity following genetic deletions of the pair, allowed us to hypothesize on the crucial role this AlcOx-Perox pair in plant penetration leading to infection.

This work is submitted for publication “Tandem fungal oxidases gate pathogenesis in major food crops”, by Bastien Bissaro, Sayo Kodama, Hayat Hage, David Ribeaucourt, Mireille Haon, Sacha Grisel, A. Jalila Simaan, Stephanie M. Forget, Harry Brumer, Marie-Noëlle Rosso, Richard O´Connell, Mickaël Lafond, Yasuyuki Kubo and Jean-Guy Berrin.

ANNEXE 3: Integrative visual omics of the white-rot fungus *Polyporus brumalis* exposes the biotechnological potential of its oxidative enzymes for delignifying raw plant biomass.

Summary of my contribution to this work

In this study, we investigated the abilities of the wood decay fungus *Polyporus brumalis* to delignify raw lignocellulosic biomass using a recent tool aimed at genome-wide visualisation of transcriptomics and secretomics. We used wheat straw as a substrate because it is considered as a model biomass and a major feedstock for bioconversion into bioenergy. We showed that the adaptive response of the fungus to wheat straw involves the recruitment of an extensive panel of oxidative enzymes. We identified co-functioning genes and co-secreted proteins contributing to the deconstruction of raw wheat straw.

My contribution to this study was the prediction of the function of genes from the CAZy family AA3_2. For this purpose, I did a phylogenetic analysis the AA3 predicted proteins from *Polyporus brumalis* and all biochemically characterized fungal AA3 genes.

Besides the set of cellulolytic enzymes commonly found in white-rot fungi, we observed in the genome of *P. brumalis* the expansion of the Class II peroxidase gene family (AA2, 19 gene copies), including manganese peroxidases (MnPs) and versatile peroxidases (VPs) and GMC oxidoreductases/dehydrogenases (AA3, 36 genes) which assist lignin breakdown by generating H_2O_2 or by reducing the oxidation products of lignin. Strinkingly, we identified 11 AA2s (five MnPs and six VPs) that were highly expressed and secreted during the fermentation. To the best of our knowledge, such an orchestrated enzymatic delignification system is unprecedented.

Our findings showed that *P. brumalis*, in belonging to the core polyporoids within Polyporales, challenges the view that the phlebioids contain the strongest lignolytic enzymatic arsenal in their genomes. Also, our findings highlighted the diversity of wood decay mechanisms present in Polyporales and the potentiality of exploring this taxonomic order for enzymatic functions of biotechnological interest.

This work is published "Integrative visual omics of the white-rot fungus *Polyporus brumalis* exposes the biotechnological potential of its oxidative enzymes for delignifying raw plant

biomass", by Shingo Miyauchi, Anaïs Rancon, Elodie Drula, Delphine Chaduli, Anne Favel, Sacha Grisel, Bernard Henrissat, Isabelle Herpoël-Gimbert, Francisco J. Ruiz-Dueñas, Didier Chevret, Matthieu Hainaut, Junyan Lin, Mei Wang, Jasmyn Pangilinan, Anna Lipzen, Laurence Lesage-Meessen, David Navarro, Robert Riley, Igor V. Grigoriev, Simeng Zhou, Sana Raouche and Marie-Noëlle Rosso, 2018, Biotechnol Biofuels 11, 201. https://doi.org/10.1186/s13068-018-1198-5